Complied for the Utah Department of Education, 2006
Updated 2007

Reading Development & Assessment of Early Literacy:

A Review of the Literature

Prepared for the Utah Department of Education
by Karin K. Hess, Ed. D. Center for Assessment, Dover, NH

April 2006
Updated February 2007

"That we can simultaneously understand reading (at multiple levels) – decoding the words, comprehending the message, and understanding the political and social context of the message – is fairly amazing. … Where reading education gets in trouble is when we try to make things too simple. We say that reading is only ___ (you can fill in the blank); then we get into trouble because we are surely leaving something out. And when we leave something out, then children do not achieve as well as when we teach them all of reading" (Stahl, K. 2005).

Purpose of this Review

The purpose of this literature review is to provide background knowledge and a research base for what is currently known about literacy development, acquisition of reading skills, early literacy assessment, and large-scale assessment as it relates to the early grades. Afflerbach (2004) states in the National Reading Conference Policy Brief that stakeholders should regularly assess the assessment to determine if it is valid (aligned with curriculum and the construct of reading), up to date (informed by the most recent research on reading and assessment) and useful (providing good information for the varied audiences of reading assessment). Information contained in this paper, along with data of current assessment practices, will be used to inform decisions about establishing early reading assessment policy and practices in the state of Utah.

With the thousands of research studies in reading available for review, it would be impossible for this paper to be comprehensive enough to include them all. Therefore, the research findings shared are those that seem most relevant to Utah's discussion of early reading assessment. Seminal research studies that have laid the foundation for current instructional practices and understanding of literacy development and major research findings that have changed our thinking about literacy learning and literacy assessment in the past decade are the focus of this review.

The primary sources for this paper were major books and national reports that have reviewed a plethora of research on reading development, reading instruction, and literacy assessment. Each of these publications makes a significant contribution to the improvement of literacy teaching, learning, and assessment, synthesizing effective research-based practices and reporting implications and implementation strategies for creating literacy-rich experiences in the early childhood classroom. Some of these resources are referenced in the introduction. All resources cited are listed in the Bibliography in Appendix D or at the end of Appendix A (Increasing Text Complexity), Appendix B (Reading Fluency Rates), or Appendix C (Menu of Frequently-Used Literacy Assessments for Grades K-3).

Topics and Sources for this Review

Best Practices in Literacy Assessment

For years, early childhood educators have effectively observed and recorded children's behavior for the purpose of chronicling development and tailoring programs to meet their rapidly changing needs. Recently, however, there has been an increase in formal testing, the results of which have been used to make "high stakes" decisions. In many cases, the instruments developed for one purpose or for a particular age group have been misapplied,

resulting in decisions that did not benefit children or their learning programs (Shepard, Kagan, & Wurtz, p. 4). In the last section of this review, Best Practices in Literacy Assessment, the sources below helped to summarize the issues and best practices for early literacy/reading assessment:

- "High Stakes Testing and Reading Assessment," a National Reading Conference Policy Brief (Afflerbach, September 2004). Presentation at the University of Maryland outlining issues, concerns, and assumptions about high-stakes testing of reading.
- *Principles and Recommendations for Early Childhood Assessments*, a report submitted to the National Education Goals Panel (Shepard, Kagan, & Wurtz, 1998), outlining principles and practices for the assessment of young children;
- *Assessment for Reading Instruction* (McKenna & Stahl, 2003);
- "Current Research in Reading/Language Arts: Understanding and Supporting Comprehension Development in the Elementary and Middle Grades" (Lipson & Cooper, 2002);
- *Early Reading Assessment: A Practitioner's Handbook* (Rathvon, 2004); and
- RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Recommendations by a national study group, headed by Catherine Snow, related to reading assessment and high-stakes testing.

Five Domains Central to Reading Instruction

In 1997, the U.S. Congress commissioned the Director of the National Institute of Child Health and Human Development (NICHD) of the National Institutes of Health, in consultation with the Secretary of Education, to create a National Reading Panel to extract the scientific research-based findings that have the greatest potential of impacting early childhood teaching practices (*Report of the National Reading Panel*, p. 3). The Panel, comprised of parents, teachers, administrators, researchers, administrators, researchers, policy makers, and other education and child development leaders, conducted a comprehensive evidenced-based review of research on how children learn to read.

The National Reading Panel reviewed more than 100,000 research studies conducted during the three decades since 1966 and 15,000 conducted prior to 1996. The major question to be answered by the panel was, "How do children learn to read?" The panel's findings help to clarify years of research in literacy development and are (selectively) addressed in this literature review, primarily under the heading of Five Domains Central to Reading Instruction (Stahl, 2005): Phonological Awareness, Phonics, Fluency, Comprehension, and Vocabulary. Additional resources consulted for research related to these 5 domains central to reading include:

- *Comprehension Instruction: Research-Based Best Practices* (Block & Pressley, 2002);
- *Handbook of Reading Research,* Volume III (Kamil, Mosenthal, Pearson, & Barr, 2000);
- *Preventing Reading Difficulties in Young Children* (Snow, Burns, & Griffin, 1998);
- "Current Research in Reading/Language Arts: Understanding and Supporting Comprehension Development in the Elementary and Middle Grades" (Lipson & Cooper, 2003);

- RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Recommendations by a national study group, headed by Catherine Snow, related to reading assessment and high-stakes testing.

Listening/Listening Comprehension and **Texts and Their Relationship to Reading Instruction and Reading Assessment**

In addition to early literacy assessment and the five domains of reading, two broad topics included for this review are Listening/Listening Comprehension (primarily as it relates to reading comprehension) and Texts and Their Relationship to Reading Instruction and Reading Assessment. Because "reading is thinking cued by written language," text types and their features are important to both reading instruction and reading assessment. Effective readers think *within the text*, picking up basic information/literal meanings; think *beyond the text*, drawing upon their own knowledge to make sense of what's read and inferring author's intent; and think *about the text*, noticing how it is crafted, appreciating the language, or understanding/critiquing the structure and message of the text (Scharer, Pinnell, Lyons, & Fountas, p.25).

Reading Strategies

Finally, the use of reading strategies (e.g., making and checking predictions, self-monitoring, using prior knowledge, etc.) is not specifically addressed in this paper, but is frequently referenced in relation to the development of reading and its importance to reading instruction: Research tells us that students learn best when teachers employ a variety of strategies to model and demonstrate reading knowledge, strategy, and skills (Braunger & Lewis, p. 94).

A reminder when reviewing research – *Reading and reading development are incredibly complex as shown by research in linguistics, cognitive and developmental psychology, anthropology, sociology, and education. "Correlation" means that two things are observed to occur together; it says nothing about whether one causes the other - or whether they interact to support each other's development (Weaver et al., 1996; Moustafa, 1995, 1997).* ***Correlation is not causation****; thus interpretations of research data must be careful in their generalizations (Braunger & Lewis, pp. 79).*

Introduction

Americans want and need good information on the well being of young children. Parents want to know if their children will be ready for school. Teachers and administrators want to know if their programs are effective and if they are providing the right programs and services. Policymakers want to know which program policies and expenditures will help children and their families and whether they are effective over time. Yet young children are notoriously difficult to assess accurately, and well-intended testing efforts in the past have done unintended harm (*Principles and Recommendations for Early Childhood Assessments,* Shepard, Kagan, & Wurtz, p. 3).

Because young children learn in ways and at rates different from older children and adults, we must tailor our assessments accordingly. What works for older children will not work for younger children. Young children (defined as age 3 to age 8/end of grade 2) have unique needs (Shepard, Kagan, & Wurtz, p. 4) that must be considered when designing assessments:

- Because they often represent their knowledge better by showing than by talking or writing, paper-an-pencil tests are not adequate.
- Because they do not have the experience to understand what the goals of formal testing are, testing interactions may be difficult or impossible to structure.
- Because they develop and learn so quickly, a test at a given point in time may not give a complete picture of learning.
- Because young children's achievements at any point are the result of a complex mix of their ability to learn and past learning opportunities, it is a mistake to interpret measures of past learning as evidence of what could be learned.

Overview of Areas of Research Reviewed

Executive Summary of Findings

1. **Listening comprehension** is not the same as reading comprehension, although they are closely related. Listening comprehension is an effective instructional strategy for remediation activities (e.g., to develop vocabulary and background knowledge) and for use when text difficulty exceeds the independent reading level of students. It can also be used to identify potential decoding problems, such as when a student's reading comprehension is compared to his/her listening comprehension of the same text.

2. **There are 5 domains essential to development of early reading** – phonological awareness, phonics, fluency, comprehension, and vocabulary - and they are interrelated in terms of development, instruction, and assessment. These domains need to be taught through a "balanced" approach – explicit instruction with frequent and regular monitoring in order to create a full picture of a student's progress in reading. Good instruction includes teacher modeling of use of reading strategies. Use of strategies, flexibly and when needed while reading, can improve a student's development of reading, especially comprehension and vocabulary. An excellent way to track literacy progress grade-to-grade is to develop a literacy profile for each student, documenting ongoing assessment results across the 5 domains of reading, as well as assessments of oral language development, concepts of print, reading strategies, reading attitudes/interests, and writing skills. (Recommended Reading Fluency Rates are included in Appendix B.)

3. **There will always be a percent of the population who have reading difficulties** and many factors outside of school can contribute to these (e.g., socioeconomic, cultural, intellectual, language barriers, etc.). Research suggests that approximately 75% of children in the bottom quartile can be brought into range of their average peers when intensive intervention (e.g., focused instruction in small groups) is implemented at the earliest time possible to address learning/reading difficulties. While more research is needed in the area of intervention, it is estimated that the lowest 10% of the entire population appears to require even further support beyond these early interventions.

4. **Text type, text features, and text complexity** are important to reading instruction and assessment. Students who comprehend narrative texts often have more trouble comprehending informational text. In low content knowledge situations, processing may be more driven, with readers relying on cues in text to organize and relate the information to achieve the intended meaning. More attention should be paid to including informational texts for direct instruction and assessment purposes. Use of increasingly difficult texts (such as leveled books) and explicit instruction in text features and characteristics of genres are essential to reading development, even at the Kindergarten level. (A Discussion of Increasing Text Complexity is included in Appendix A.)

5. **Assessment of early literacy** should include evidence collected for all 5 domains of reading, along with assessment of use of reading strategies and habits of reading to create a complete picture of student learning. Three of these 5 domains of reading - phonological awareness, phonics, and fluency - are best assessed individually (in the classroom) than on a large-scale assessment. There is evidence to suggest that interviewing students (e.g., using questioning, asking them to retell a story) might be a more effective method of assessing comprehension

of texts at the early grades. Systematic observation, under controlled conditions, is by far the most effective assessment strategy recommended for early literacy (such as use of Clay's *Observation Survey*). Individually administered tests of reading tend to provide more dependable results than group assessments, because the teacher can command the mental engagement of the student, to some extent, during the testing process. (A Menu of Frequently-Used Literacy Assessments for monitoring the 5 Domains of Reading is included in Appendix C.)

6. **An integrated system of reading assessments – not simply one assessment,** is recommended by many national groups (Rand Reading Study Group, 2002; National Reading Council, 2004). This assessment system should include a variety of assessments appropriate for different purposes and specific groups (e.g., emergent or beginning readers, older struggling readers, second-language readers). Further, the various assessments included in the system would address different purposes, such as a portmanteau assessment for accountability or screening purposes, diagnostic assessments for guiding intervention, curriculum-linked assessments for guiding instruction, and so on.

7. **A variety of group-administered standardized measures have been developed** to assess literacy outcomes. They include test like the *Stanford Achievement Tests*, *California Achievement Tests*, *Iowa Tests of Basic Skills*, *Terra Nova*, *Gates McGinitie Reading Tests*, etc. While group tests are far more efficient and cost effective than individually administered assessments, their results are less dependable, especially at the earlier grade levels.

8. **The public, in general, supports high stakes testing and believe that the tests are fair and "scientific."** These tests have the ability to reduce and summarize complexities of reading to single raw scores and percentile rankings, and in doing so they appear almost magical. Large-scale assessments need to demonstrate a high level of *content* validity (reading content as described in the standards) and *construct* validity (reading ability assessed through actual reading of/or engaging with text); these technical requirements for validity may not be achievable, especially at the Kindergarten through second grade levels. The International Reading Association (1999) adopted a position paper against the use of high-stakes testing due, in part, to their tendency to narrow curriculum, especially in high poverty areas. These concerns about high-stakes testing are also voiced in the RAND Reading Study Group Report (2002) and the National Reading Conference Policy Brief (2004). There is currently no compelling evidence to support large-scale assessment of reading below grade 3.

9. **Some states are addressing early literacy assessment** in conjunction with local (classroom/school/district) assessment of literacy through the creation of literacy profiles to document progress. Several states have early literacy assessment guidelines for grades K-3 and/or approaches to large-scale assessment that require classroom teachers to administer the assessments to their students and report results. Further research could explore how these assessment approaches are working and what other states are doing *successfully* in the area of early literacy assessment.

10. Research shows that **teachers** (who will be making instructional and curricular decisions) tend not use or value test results from large-scale or standardized assessments (National Reading Conference Policy Brief, 2004). Classroom teachers tend to see greater value in formative assessment and skilled teachers do it constantly, relying heavily on classroom assessment to make instructional and placement decisions. The International Reading Association's position paper opposing high stakes literacy testing (1999) concludes by providing the following recommendations to teachers: "Construct more systematic and rigorous assessments for classrooms, so that external audiences will gain confidence in the measures that are being used and in their inherent value to inform decisions."

Research Related to Listening/Listening Comprehension

Many parallels exit between the learning of oral language and learning of print language. Reading, writing, speaking, and listening, at the deep levels of production and comprehension are parallel manifestations of the same vital human function – the mind's effort to create meaning (Cambourne, 1988). Cognitively, the same processes seem to be in effect with development of all language processes. However, the two modes of language are different in many complex and interesting ways. The most salient difference is that the two require different kinds of knowledge that learners must acquire in order to operate with and on them.

Certain uses of the written mode require specific knowledge that cannot be carried over from the oral mode, and visa versa. For example, written language uses knowledge of commas and capital letters; oral language uses pauses and intonation. Trumbull and Farr (2005) note an important cognitive difference between oral and written language – the fact that written language is doubly symbolic; readers and writers must become at least somewhat conscious of their knowledge of oral language, which is not necessary in speaking most of the time (Flood & Menyuk, 1981). (Braunger & Lewis, pp. 27-28)

The number of research studies on listening is small, and the effectiveness of listening in relation to reading comprehension lacks a strong scientific base. It is more likely that listening occurs informally as part of reading and content area instruction (National Reading Panel, p. 104, section 4). Much literacy instruction for young children comes through read-aloud and shared reading, as well as actively listening to and discussing stories. Listening and listening comprehension are frequently used at all grade levels when the complexity of text exceeds readers' instructional and independent reading levels. Assessment strategies employing listening and/or listening comprehension tend to be most effective when used with remedial students (grades 1-6) to improve language acquisition/vocabulary development or to identify deficiencies in decoding skills.

- Much literacy instruction for young children comes through read-aloud and shared reading, where the primary responsibility of processing *printed* text is in the hands of the teacher, rather than the student. More research is needed to understand how/whether guided retellings of stories read aloud transfer to their own reading later in schooling (Block & Pressley, pp. 256-257).
- Researchers have found that the following oral language skills contribute to success in formal reading instruction: large vocabulary; use of correct grammatical forms in

conversation; ability to sustain conversations; and motivation to use language to solve problems (Hall & Moats, 1999) (*PA Literacy Framework*, 2000a, p. 2.7).

- Assessments of young children's reading tend to focus on word reading rather than on comprehension. Assessments of listening comprehension and of oral language production, both of which are highly related to reading comprehension, are rare and tend not to be included in reading assessment systems despite their clear relevance. The available listening comprehension assessments for young children do not reflect children's rich oral language–processing capacities, because they reflect neither the full complexity of their sentence processing nor the domain of discourse skills (RAND, p. 56).
- Reading, in contrast to oral language, does not emerge naturally from interactions with parents and other adults, even when children are exposed to literacy-rich environments before entering school. Research findings have been unclear about the degree of explicitness and the intensity and duration of instruction, but most children require systematic and explicit instruction over a relatively long period of time to learn to read (*PA Literacy Framework*, 2000a, p. 2.19).
- One way to start comprehension instruction early (identification of story themes) is to bypass reading and work on an oral level, reading stories to children and then discussing the story with them. Research has show that with this approach (with at-risk second and third graders) it is important to choose stories with one clear, simple, and accessible theme to start with, saving more complicated stories for when students have definitely mastered the concept of theme scheme (i.e., levels of meaning going beyond the specific plot) (Block & Pressley, pp. 134-135). Further research studies indicate that theme is the most difficult story component to teach and that even with extensive teacher modeling and direct teaching, there was little success in improving a student's ability to identify story theme using only listening comprehension (Block & Pressley, pp. 128-129).
- Four research studies conducted (1984-1995) on the use of active listening instruction were aimed at improving critical listening, reading, reading comprehension, and increasing students' participation in discussions and more thoughtful responses to questions. Generally, the teacher guided the students to listen while others read, then posed questions for the students to answer. These studies targeted readers in grades 1-6, identified as remedial or at-risk students, and showed some improvement in reading comprehension as measured by teacher questioning or standardized tests (Block & Pressley, pp. 178 and 191).
- Many basic cognitive processes are shared during reading and listening. Syntactic and inferential processes, as well as background and word knowledge play a role in both. The correlations between listening and reading comprehension are high for adult populations. Studies tend to find high correlations between reading and listening comprehension *after* the child has learned to decode; however, the gap between's one's listening and one's reading comprehension can in fact be quite large even when correlations between the two are quite strong (Snow, Burns, & Griffin, pp. 64-65).
- A child's listening level is conventionally defined as the highest passage at which comprehension of the text read aloud to the child is at least 75% comprehension. Generally, the teacher estimates the student's reading frustration level and then selects the next-higher passage to administer on a listening basis. Both the passage and

questions are read aloud. Knowing a child's listening level can be useful in discerning whether comprehension difficulties are the result of decoding problems. For example, a fourth grader could have a listening level of 4th grade, but a reading comprehension level of only second grade, due to inadequate word-recognition skills (McKenna & Stahl, pp. 60-61). Listening to texts read aloud is often used as an instructional strategy when the text difficulty exceeds the child's reading skills.

- If comprehension improves when a child listens to a selection read aloud, the improvement would be evidence of inadequate decoding ability as the fundamental problem; however, often a child will need direct instruction in decoding and comprehension, and perhaps oral language development to improve reading comprehension (McKenna & Stahl, p. 177).
- Listen-Read-Discuss is a reading strategy introduced by Manzo and Casale (1985) to permit students possessing too limited knowledge/experience and breadth of vocabulary to be able to fully comprehend material during the first reading of it. Teachers present all of the content covered in the reading selection prior to students reading the text. This has been a proven a viable instructional approach when students have too limited a vocabulary to comprehend informational text (McKenna & Stahl, p. 178).
- "Read-alouds" (an instructional strategy of reading texts aloud to students as they follow along) assist all students with language acquisition by enabling them to become familiar with the academic or literary language necessary for school success (Se′ne′chal, Thomas, & Monker, 1995). For example, research shows that second language learners acquire vocabulary and grammar from read-alouds, and read-alouds can improve young students' comprehension of decontextualized language (Beck & McKeown, 2001). Kindergarten students who have been read to are more likely to emerge as readers earlier than those who did not often hear stories (Sulzby, 1985). (Meyers, p. 315)
- Assessing sight-word knowledge (using lists of isolated words, not in text) can be accomplished through group assessment, using the format of presenting students with row after row of 4 word choices. The teacher reads a word aloud and instructs students to circle the correct word read in each row. The advantage of this approach is efficiency (group assessment); however, the accuracy of results may be comprised when letter cues are used to eliminate possible responses without students actually reading the words. More accurate results are achieved when individual students read words orally when presented to them (such as with flash cards) (McKenna & Stahl, pp.111-112).
- There are several problems associated with the use of listening comprehension as a predictor of reading ability: (1) there are still very few listening comprehension measures with adequate psychometric characteristics; (2) as on IQ tests, children with limited English proficiency may score poorly on measures of listening comprehension because of difficulty understanding test directions; and (3) there is some evidence that the listening comprehension-reading comprehension discrepancy model may over identify children as reading disabled (Rathvon, pp. 106-107).

Research Related to Domains Central to Reading Instruction: "The Big 5"

The National Reading Panel (2000) determined that "effective reading instruction includes teaching children to break apart and manipulate the sounds in words (phonemic awareness), teaching them that these sounds are represented by letters of the alphabet which can then be blended together to form words (phonics), having them practice what they've learned by reading aloud with guidance (guided oral reading), and applying reading comprehension strategies to guide and improve reading comprehension."

Five domains of reading identified by the National Reading Panel as essential to instruction and assessment of early reading are: Phonological Awareness, Phonics, Fluency, Comprehension of Text, and Vocabulary. While all 5 of these areas may not be assessed on a state-level (large-scale) assessment, assessment evidence across the 5 domains should be used together within the (local) district assessment system to paint a more accurate picture of students' progress and needs in reading. Policy recommendations for state assessment need to consider the complimentary nature of state, district, and classroom level assessments.

Assessment evidence (of the "Big 5" domains of reading) at different levels of system should compliment each other in the following ways:

- Provide related information using multiple, assessment formats;
- Provide differing levels of specificity (see table below);
- Provide varied applications of content knowledge and skills;
- Provide information at different points in the student's schooling;
- Support decisions for teaching and learning; and
- Provide information valued at the local level.

Assessing Progress Towards Achieving Standards: Levels of Specificity (Source: Karin Hess, presentation at Vermont Literacy Institute, Killington, VT, 2004)	
State Level Assessment	Used to benchmark student performance at specified points related to achievement of standards (e.g., comprehension); must be well-aligned to constructs identified in standards (actually assess identified skills)
District Level /School Level/ Grade Level Assessment	Used to assess specific content and skills that students are required to master by that grade level; often provide more depth or breadth of evidence than state assessments can; include periodic assessments used to inform instructional programs, curriculum, and student placement decisions
Classroom Level Assessment	Are generally more specific to content and skills, more frequently administered, and more narrowly targeted than periodic school level or state level assessments; May measure prerequisite skills, skills from earlier or later grade levels, or mastery of skills and concepts; include ongoing formative assessments used to inform instructional decisions

Selected research findings related to each of the 5 domains central to reading are included on the following pages. In many cases, there will be overlap of the categories due to the interrelated nature of the domains of reading. For example, expressive language of young children (vocabulary) requires accurate retrieval of stored phonological representations (Snow, Burns, & Griffin, p. 109); and fluent reading affects ability to comprehend text

(McKenna & Stahl, p.12). A research project on intervention strategies addressing oral reading of text, found that the effect for comprehension was almost as substantial as the effect for vocabulary and fluency, providing further confirmation of the strong interrelatedness among the components of reading (Eldredge, Reutzel, & Hollingsworth, 1996) (Kamil et al., pp. 470-471).

NOTE: The five domains central to reading instruction included in this section of the literature review comprise a slightly different "list" than those that the Reading First initiative focuses upon.

Reading First Outcomes to be Assessed at Each Grade Level, K-3 (Torgesen, Kame'enui, Francis, Fuchs, Good, O'Connor, Simmons, & Tindal, 2002)	
Kindergarten	Phonemic Awareness, Phonics, and Vocabulary
1st Grade	Phonemic Awareness, Phonics, Fluency, Reading Comprehension, and Vocabulary
2nd Grade	Fluency, Reading Comprehension, and Vocabulary
3rd Grade	Fluency, Reading Comprehension, and Vocabulary

Phonological Awareness (*which includes Phonemic Awareness*)

It is important to point out a few distinctions between and among the terminology used in this section. Phonemic awareness is not phonics. Phonemic awareness is the understanding that the sounds of spoken language work together to make words. Phonemic awareness is a subcategory of phonological awareness and has a more narrow focus – identifying and manipulating the individual sounds in words. Phonological awareness includes identifying and manipulating larger parts of spoken language, such as base words and syllables (Armbruster & Osborn, pp. 2-4).

- Phonemic awareness instruction teaches children to notice sounds in words and improves a student's ability to read words, to comprehend, and to learn to spell (Armbruster & Osborn, pp. 5-6).
- On average, phonological awareness has been shown (24 studies examined) to be about as strong a predictor (r = .46) of future reading as memory for sentences and stories and general language development. [Note that r = .46 would be considered a "modest" correlation.] Phonological awareness in kindergarten appears to be a more successful predictor of future *superior* reading than of reading *problems* (Snow, Burns, & Griffin, p. 112).
- The extent to which children lack phonemic awareness, they are unable to internalize their phonics lessons. The result is difficulty in sounding and blending new words, retaining words from one encounter to the next, and learning to spell. Research repeatedly demonstrates that with normal, as well as at-risk populations, adequate awareness of phonemes accelerates reading and spelling growth (Snow, Burns, & Griffin, pp. 55-56).
- Phonological decoding is a routine part of skilled word identification (Snow, Burns, & Griffin, p. 65).
- One of the most striking findings of a 1988 study that followed 54 children from kindergarten through to grade 4 was that there was an 88% chance that if a child was

having difficulty reading at grade 1, that child would still have difficulty with reading at grade 4. More importantly, the best predictor of poor reading achievement at grade 1 was low phonemic awareness. Later studies have shown that readers of all ages who lack phonemic awareness show difficulty in reading and spelling (Pressley, pp. 106-107).

- Phonemic awareness is a better predictor of reading achievement than more global measures of general intelligence (I.Q. tests) or reading "readiness" (Adams, 1990); Blachman, 1989, 1991; Catts, 1991; Griffith & Olson, 1992; Stanovich, 1986; Yopp, 1995). Developing phonemic awareness requires the advancement through eight types of tasks that have a positive effect on reading acquisition and spelling (e.g., rhyming, auditorily discriminating sounds that are different, blending spoken sounds into words, word-to-word matching, isolating sounds in words, etc.). Research findings indicate that educators should begin teaching phonemic awareness directly in kindergarten (or even earlier) (*PA Literacy Framework*, 2000a, p. 2.23).
- Phonemic awareness instruction can account for about 12% of word recognition performance in the early primary years; although it explains much less of the variance in later primary and middle grades, the effects are still detectable (Pressley, p. 117).

Phonics

Phonics is the understanding that there is a predictable relationship between phonemes (sounds) and graphemes (letters that represent sounds in written language) (Armbruster & Osborn, pp. 2-4). A wide variety of methods can be used to teach phonics – intensive, explicit, synthetic, analytic, and embedded instruction. All instructional methods focus the learner's attention to the relationship between sounds and symbols as a strategy for word recognition. There continues to be insufficient evidence that one form of phonics instruction is strongly superior to another (Allington, 1997; Cunningham & Allington, 2003; Stahl, McKenna, & Pagnucco, 1994) (Braunger & Lewis, p. 84).

- Systematic and explicit phonics instruction significantly improves kindergarten and first grade students' word recognition, spelling, and reading comprehension. Phonics instruction is most effective when begun early (kindergarten or first grade) but should not be seen as the entire reading program for beginning readers. Beginning readers should be applying letter-sound relationships (phonics), and listening to, and reading stories and informational texts aloud and silently (Armbruster & Osborn, pp. 13-19).
- Numerous studies over more than forty years have looked at the effects of systematic phonics instruction (explicitly taught letter-sound associations, such as short vowel sounds, blending sounds, etc.), analytic/implicit phonics (analyzing sounds of known sight words and words used in context to decode words), phonograms (using chunks of letters – e.g., *ill, ick* - to decode unknown words), and analogies (comparing one word to another – e.g., comparing *strike* to *like*). Studies have found that children need to be able to understand letter-by-letter decoding before they are able to benefit from phonogram and analogy instruction, but children need to use phonograms in order to read proficiently. Children first use phonograms and analogies to decode monosyllabic words; later they can use them to decode polysyllabic words (McKenna & Stahl, p.13-14).

- Children need to know how patterns (not rules, such as "when 2 vowels go walking") help them to identify unknown words (McKenna & Stahl, p.14).
- According to the NRP report (NICHD, 2000), the benefits of phonics instruction are strongest in first grade and are diminished for students in subsequent grades. Older readers who are identified as "struggling" lack experiences in reading extended coherent texts (Allington, 2001; Schoenbach, Greenleaf, Cziko, & Hurtz, 1999) and need instruction in ways to construct meaning with text rather than instruction in phonics (Braunger & Lewis, p. 85).

Fluency

When fluent readers read silently, they recognize words automatically, quickly group words to gain meaning, and read with expression when reading orally. Repeated and monitored oral reading improves fluency and overall reading achievement. Readers who have not yet attained fluency will not likely make effective use of silent/independent reading time (Armbruster & Osborn, pp. 22-30). Appendix B includes a synthesis of the current literature with recommended silent and oral reading fluency rates found in.

- Isolated word recognition is a necessary, but not sufficient condition, for fluent reading (Armbruster & Osborn, p.30).
- Competent reading requires that skills extend beyond the single-word level to contextual reading, and this skill can be best acquired by practicing reading in which the words are in a meaningful context (National Reading Panel, p. 11, section 3).
- Because fluent readers do not have to concentrate on decoding the words, they can focus their attention on what the text means (Armbruster & Osborn, p. 22).
- Many children with reading problems can read accurately, but cannot read quickly enough to make sense of what they are reading. If reading is not fluent, then comprehension usually suffers (McKenna & Stahl, p.12).
- Repeated reading with feedback and guidance was found to be superior to repeated reading alone for improving fluency (National Reading Panel, p. 15, section 3).
- The analysis of guided reading procedures led to the conclusion that such (instructional and formative assessment) procedures had a consistent, and positive impact on word recognition, fluency, and comprehension (National Reading Panel, p. 3, section 3).
- Preliminary results from a 2005 study, seem to indicate that it is not the repetition of text itself that is the key to the development of fluency, but the use of scaffolded supports and the focus on extensive oral reading of more difficult text that lends to the effectiveness of methods used to improve fluency (Stahl, K. p. 187).
- Students who scored lower on measures of fluency also scored lower on measures of comprehension, suggesting that fluency is a neglected skill in many American classrooms, affecting many students' reading comprehension (Armbruster & Osborn, p.23).
- It is clear that fluency may also include the ability to group words appropriately into meaningful grammatical units for interpretation, enabling reading comprehension, by freeing cognitive resources for interpretation (National Reading Panel, P.6, section 3).

Comprehension of Text

True reading comprehension and subsequent reading engagement requires more than cognition; it means entering textual worlds, maintaining a balance between engrossment and critical distance, and formulating one's own response to various dilemmas in text (DiPardo & Schnack, 2004). Research has shown that many children who read at the third grade level in grade 3 will not automatically become proficient readers in later grades unless teachers explicitly teach strategies for reading comprehension, especially for poor comprehenders (Bishop, Reyes, & Pflaum, 2006, p. 66).

Comprehension is the *reason* for reading – a cognitive activity relying on excellent fluency, vocabulary, and prior knowledge. Reading is not a passive activity; active interactive processes are critically necessary to the development of reading comprehension (National Reading Panel, p. 11, section 4); and even able readers benefit from explicit instruction in the use of reading strategies (National Reading Panel, p. 47, section 4). All reading assessment must be clearly and carefully tied to an informed understanding of what reading "is." We are fortunate to have a rich knowledge base that describes the nature of reading and its development (Clay, 1979; Heath, 1983; Snow, 2002). From this research we are able to construct an understanding of reading that should guide our efforts to design and use reading assessment effectively (Afflerbach, 2004, pp.2-3).

- Rosenblatt's (1978) transactional theory of literature defines reading as an interactive transaction between the reader and the text. During the comprehending process, the reader creates personal meaning inspired the text. Meaning resides within the exchange and not in the text alone. Rosenblatt (1978, 2005) made a distinction between responding from an aesthetic and an efferent stance or point of view. At one end of the continuum are readers' personal experiences, feelings, and emotions; at the other end are responses that reflect attention to text features, such as identifying facts or text structure (Heller, p. 359; RAND, p. 32).
- Research supports the effectiveness of 6 reading strategies that improve comprehension: (1) student self-monitoring of reading (e.g., knowing when they do not understand and then self-correcting), (2) using graphic organizers to illustrate relationships and ideas, (3) answering questions (explicit and implicit in text, and using prior knowledge), (4) generating questions to process information, (5) recognizing story structure, and (6) summarizing central ideas (Armbruster & Osborn, pp. 48-53).
- Many students experience comprehension problems with expository text. There are several reasons for this, one being that they cannot see the basic internal structure of text. Students with a good understanding of text structure have fewer comprehension problems. Teaching expository text structure explicitly and systematically will improve reading comprehension (Dymock, pp. 171-178).
- Research (1999, 2000) of exemplary first grade teachers indicates that direct instruction in comprehension strategies, along with a strong word-analysis program (decoding), provided a multitude of opportunities for reading and writing high-quality texts. In a study of exemplary teachers and schools serving high poverty populations (2000), integrating direct instruction of comprehension with decoding skills at the

primary grades led to more higher-order questioning about texts. Comprehension instruction included: story elements (setting, character, etc.), story retelling, summarizing, interpreting, predicting, discussing relationships, and text structure analysis (Block & Pressley, pp. 251-254).

- In a series of studies with 7 and 8-year olds (1991 and 1996) comparing skilled and non-skilled comprehenders, skilled comprehenders demonstrated greater understanding of pronoun referents, made proper inferences about the texts from particular words, drew more global inferences from elements of the texts that were not adjacent, detected inconsistencies in texts, applied background knowledge, retold stories with events more integrated, and monitored their own comprehension (Snow, Burns, & Griffin, pp. 76-77).
- Because early reading instruction emphasizes word recognition rather than comprehension (Cain, 1996), the less skilled comprehenders' difficulties generally go unnoticed by classroom teachers (Snow, Burns, & Griffin, p. 77).
- There has been a great deal of testing of reading comprehension (Pressley, 1998), but little systematic teaching of reading comprehension strategies. When teachers applied strategy instruction in the classroom, even when they omitted crucial aspects of strategy use, their students improved in reading comprehension (Block & Pressley, pp.186-187).
- Even kindergarten-age children's comprehension can be improved through guided comprehension instruction. When one group of students were asked to draw about and another group asked to retell stories heard read aloud, after 8 weeks of instruction the guided-retelling group performed substantially better on both comprehension and retelling than the group who only drew pictures to demonstrate comprehension (Block & Pressley, p.248).
- Comprehension may be poor with the first reading of text, but with each additional reading, the student is better able to comprehend, because the decoding barrier to comprehension is gradually overcome. As less attention is required for decoding, more attention is available for comprehension. Thus rereading builds both fluency and comprehension (Samuels, p. 405).
- Powerful comprehension and profound enjoyment of texts comes only when readers appreciate and use an array of devices employed by authors to build meaning. The techniques and devices (collectively called author's craft) are critical for comprehension. When children make good use of author's craft (e.g., exaggeration, dialogue, flashback, imagery, theme), their comprehension and enjoyment are enhanced. When they do not, they frequently misunderstand or fail altogether to assign meaning to the events and action of a story (Lipson & Cooper, p. 9).
- *Best practices in monitoring development of reading comprehension* that are supported by research include: responding to text (using written responses, role playing, discussion, book talks); retelling (both orally and in writing); individually administered reading inventories; interviews; work samples (e.g., reviewing reading logs, *Meisels' Work Sampling System*); conferencing with individual students; teacher observation; teacher anecdotal records; and student self-assessment (e.g., applying a comprehension rubric to a specific task) (Braunger & Lewis, p. 135).

- Evidence suggests that comprehension is facilitated if there is congruence between the structure of text and the structure of the conceptual domain (Zwaan & Radvansky, 1998). (Kamil et al., p. 318)
- Children in the first years of school – kindergarten and first and second grade – exhibit several conceptual types of resistance to stories: intertextual (between known text and new text); preferential or categorical (what a child likes to read); reality testing (text does not reflect the child's world); engaged or kinetic (too painful a subject); exclusionary (can't relate to story); and literary critical (perceived faults in author's craft)."Expressions of resistance are expressions of active readers engaged in the construction of meaning from texts and life experiences. … Literary critical resistance suggests the emergence of a literal sensibility in children as they become increasingly able to discuss the merits of particular books. These comments indicate children's awareness of narrative elements (plot, setting, character, and theme) and the functions they perform in the construction of a coherent story" (Sipe & McGuire, 2006, pp. 7-12).
- Consistent questioning that encourages recitation in reading (based on factual recall only) represents missed opportunities for teachers to model for their students the true nature of reading. "As Ruddell (2001) powerfully emphasizes, the questions teacher ask help to shape student understanding of and expectations about reading comprehension." Similar sentiments have been echoed by reading researchers and theorists for many years (Dickenson & Smith, 1994; Pearson & Fielding, 1991; Langer, 1992). Teachers who regularly use higher order questions to engage pupils in discussion are likely to find that that their students read more (Guthrie, Schaefer, Wang, & Afflerbach, 1995) (Applegate, Quinn, & Applegate, p. 48).
- In grades 3–5, engaging students in elaborative questioning improves their comprehension of text read during instruction and their comprehension of new text read independently. Similarly, teaching students in grades 3–9 to self-question while reading text enhances their understanding of the text used in the instruction and improves their comprehension of new text. Teaching students in grades 3–6 to identify and represent story structure improves their comprehension of the story they have read. In the case of this strategy, there was no evidence that the strategy transferred to the reading of new stories and improvement was more marked for low-achieving readers (RAND, p. 33).
- Research with poor comprehenders has been motivated by a particular set of hypotheses about impediments to comprehension. Some of these hypotheses suggest that the problems of poor comprehenders are an outgrowth of differential instruction; that is, these students have been denied the kinds of instruction that advance reading comprehension. This hypothesis is particularly relevant for students who have a history of reading problems (e.g., decoding problems in grades 1 and 2). For example, McDermott and Varenne (1995) documented that teachers working with high-achieving students focused on higher-order thinking with text and communicated clearly that the purpose of reading is understanding. In contrast, these same teachers, when working with low-achieving students, focused on low-level factual reading, interrupted children's reading more frequently than their errors would justify (see also Shake, 1986), and communicated little about comprehension as the goal of reading. A corollary to this hypothesis is that students with a history of reading challenges read

less text; hence, they accrue less background knowledge to bring to the reading of new text (RAND, p. 34).

- Assessments that target particular operations involved in reading comprehension must be able to identify subtypes of poor comprehenders in terms of the components and desired outcomes of comprehension. They should be capable of identifying both intra- and inter-individual differences in acquiring the knowledge and skills necessary for becoming a good comprehender (RAND, pp. 56-57).

Vocabulary

Research on literacy learning clearly shows that processes of reading, writing, speaking, listening, viewing, and thinking develop simultaneously as learners become literate (Cooper, 2000). Language arts methods recommended today capitalize on the fact that that all of these aspects develop together, yet need to address children with varied schemas. As teachers encounter more and more students with diverse backgrounds, languages, and educational experiences (Short, Echevarria, 2004), they need to employ logical instructional approaches that build common backgrounds or schema about a topic and broaden students' vocabulary (Dorr, pp. 139-140).

Although a great deal of vocabulary is learned indirectly, direct instruction in vocabulary will help students learn the meaning of specific words and more importantly, to develop word-learning strategies, such as meanings of word parts (affixes, base words), and use of context clues and dictionaries to unlock meaning of unknown words. Vocabulary development directly affects reading comprehension (Armbruster & Osborn, pp. 35-40).

- Several kinds of vocabulary measures have been used as predictors of future reading achievement after kindergarten. In a review of twenty studies of receptive language (the child indicates which of several pictures corresponds to the word spoken by the examiner), the mean correlation between receptive vocabulary scores in kindergarten and subsequent reading in grades 1-3 was only r =.36. Compared with receptive tests, measures of expressive language, which place greater demands on accurate retrieval of stored phonological representations, the average correlation with future reading ability is r = .45. The consistency of findings of five kindergarten prediction studies strongly suggests that expressive language/naming vocabulary (accuracy and speed of object naming) is a reliable predictor of future reading ability (Snow, Burns, & Griffin, pp. 109-110). [Note that r = .45 would be considered a "modest" correlation.]
- Comprehension research has clearly demonstrated the importance of the reader's background knowledge for understanding texts, such as how knowledge of content plays an important role in formulation of main ideas. Vocabulary knowledge has long been known to be a correlate of comprehension ability as measured on standardized tests (Snow, Burns, & Griffin, pp. 62-63).
- Stanovich (1986) suggests that many of the problems struggling readers encounter are not due to underlying causes, but to these children's increasing lack of reading experiences. In the case of vocabulary, children with reading problems both read fewer texts and read less challenging texts. They are therefore exposed to fewer words of increasing difficulty, and because words are learned from exposure in context, they learn` fewer words (McKenna & Stahl, pp.117).

- The breadth and depth of a child's literacy experiences determine not only how many and what kinds of words s/he will encounter, but also the background knowledge with which a child can conceptualize the meaning of new words. The ultimate significance and memorability of any word or text depends upon whether children possess the background knowledge and conceptual sophistication to understand its meaning (Snow, Burns, & Griffin, p. 219).
- Written texts place high demands on vocabulary knowledge. Even words used in children's books are more rare than those used in adult conversations and prime time television. Learning new concepts and the words that encode them is essential to comprehension development (Snow, Burns, & Griffin, p. 217).
- Studies of how children acquire vocabulary indicate that is a complex process. In their review of vocabulary processes, Nagy and Scott (2000) described the complexity of word knowledge. Factors include: word knowledge is incremental (multiple exposures over time); words have multiple meanings (and students need to *expect* that they may have more than one meaning); words are multidimensional (e.g., definitions, pronunciations, use, how it relates to other words/concepts); and word knowledge is interrelated (use of background knowledge better prepares students for new word encounters) (Spencer & Guillaume, pp. 206-207).

Research Related to Text Complexity, Reading Instruction, and Assessment

Increasing text complexity, characteristic text features, internal text structure, and text types/genre are all important to the instruction and assessment of reading. Reading assessment should reflect performance over multiple time points with various texts and purposes. We must provide students with opportunities to demonstrate their reading growth and achievement in situations that reflect their daily lives as readers (Afflerbach, 2004, pp. 11-15).

For many years, only narrative literature was used for reading instruction with young children. Today, the importance of exposing students of all ages to a wider range of texts and their characteristic features is getting more and more attention. It is estimated that as much as 80% of what students will be expected to read throughout their lives will be nonfiction. Appendix A of this paper includes a compilation of descriptors of texts recommended for each grade level or grade span, as identified in the literature.

- The body of research on young children's comprehension development is based on a limited diet of text (almost always fictional narrative). Yet there is growing evidence that children should be provided with greater diversity of texts in early schooling; the literacy demands posed by the information age are nothing like we've seen before. There is a need for research that examines how to support children's comprehension of nonlinear and non-story texts, including informational and procedural texts. To delay this sort of powerful instruction until children have reached the intermediate grades is to deny them the very experiences that help children develop the most important of reading dispositions – the expectation that they should and can understand each and every text they read (Block & Pressley, p. 257).
- Unlike readability formulas that simply count words or syllables, an analysis of the complexity of a narrative story or expository text requires close reading and attention to relationships between and among ideas (Lipson & Cooper, p. 5).

- The real literature that teachers want to use is not written with literacy instruction in mind and often presents special challenges to emergent readers. But turning to the traditional readability formulae to guide text selection is not an answer either, because these formulae are based upon a very limited number of factors (Saul & Dieckman, p. 507).
- Research (Pappas, 1991; Pappas, 1993) examined the emergent reading of narrative stories and informational texts by 16 kindergarten students over 4 months and found that children were equally able to negotiate complex text structure differences between narrative and informational books. Related research (Smolkin, Yaden, Brown, & Hofius, 1992; Yaden, 1993; Yaden, Smolkin, & MacGillivray, 1995) found that certain features of alphabet books (e.g., illustrations, text balloons, print) changed the nature of children's responses and discussion about books, including the graphic nature of text and conventions of print (Kamil, Mosenthal, Pearson, & Barr, p. 430).
- Research identifies proficient readers of informational texts as attending to both the external physical organization and features of text (e.g., tables, graphics, glossary, headings) and internal structure of ideas (e.g., compare/contrast, description, cause/effect) and employ a small set of powerful comprehension strategies (e.g., using text clues, text features, locating and organizing information) (Block & Pressley, pp. 262 – 267 and 290-293).
- On the whole, research in teaching text structure has indicated that explicit teaching (in this case explicit teaching of particular text structures) can improve comprehension and composition of those structures (Dickson, Simmons & Kame'ennui, 1998a, 1998b; Gersten, Fuchs, Williams, & Baker, 2001). (Purcell-Gates, Duke, & Martineau, p. 11).
- Recent research suggests a scarcity of nonfiction and expository prose in the primary grades (Moss, 2004; Palmer & Stewart, 2005) (Heller, p. 358).
- Balanced attention to narrative and informational text is needed from the start. Children's ability to engage in search tasks is likely to be enhanced by balanced exposure to informational text throughout the elementary grades (Block & Pressley, p. 294).
- Students with a good understanding of text structure have fewer comprehension problems. Teaching expository text structure explicitly and systematically will improve reading comprehension (Dymock, p. 171-178).
- Research suggests that pictured events and concepts are significantly more likely to be recalled than non-pictured events (Lipson, Mosenthal, & Mekkelsen, 1999). If pictures are central to and support the main themes and ideas of the story, this is good. However, if the pictures are not supportive, or draw children's attention to unimportant side events (called "seductive details") this can pose comprehension problems (Alexander & Jetton, 2000). Older, more mature readers do not rely so heavily on pictures to comprehend texts read (Lipson & Cooper, p. 3).
- Some features of text influence comprehension for all readers. Aspects of text, such as its *structure*, *complexity*, and *genre* affect understanding even when readers are very accurate decoders (Goldman & Rakestraw, 2000; Lipson et al., 1999) (Lipson & Cooper, p. 3).
- Recent research suggests that extensive reading practice – for primary grade children, between 20-30 minutes each day; for intermediate children as much as an hour each

day is important to planned instructional programs. Effective reading programs must include ample opportunity for students to read appropriately leveled texts (Lipson & Cooper, p. 10).

- Although short, contrived texts can be helpful for introducing a reading skill or strategy, students will not be able to develop effective comprehension strategies such as self-monitoring, summarizing, and self-questioning unless they are reading increasingly complex material of appropriately substantial length. Nor will they develop and acquire the rich vocabulary and broad understanding of text structure required to become a reader with excellent comprehension (Lipson & Cooper, p. 10).
- The current emphasis on testing has generated greater attention to nonfiction text. In an analysis of reading passages on standardized tests, Calkins, Montgomery, Santman, and Falk (1998) found that 50-85% of the texts used on standardized tests are informational. Thus, many educators argue that student performance will improve if teachers attended more to a genre that is so frequently tested (Saul & Dieckman, p. 503).
- Leveled books, used today in many primary classrooms, are categorized along a gradient of difficulty to help teachers organize small-group instruction. They provide a ladder of support so students can take on more difficult texts with teacher support, and in the process, expand their strategies for thinking within, beyond, and about texts (Scharer, Pinnell, Lyons, & Fountas, p.27).
- Because so little is known about the contribution of texts to early reading instruction, much further research is needed, especially given that certain kinds of texts may be mandated without evidence of their effectiveness in supporting instruction. It seems prudent to follow the consensus of professional opinion that books for early reading instruction should be leveled, and leveled along curricular dimensions of the instructional emphasis the books are expected to support (Cunningham et al., pp. 425-426).
- Generally, in situations of high content knowledge, readers will be less reliant on structural aspects of the text than in low content knowledge situations, because they can draw on preexisting information to create accurate and coherent mental representations. In low content knowledge situations, processing may be more driven, with readers relying on cues in text to organize and relate the information and achieve intended meaning (Beck & Dole, 1992; Beck, McKeown, Sinatra, & Loxterman, 1991; Britton & Gulgoz, 1991; McNamara & Kintsch, 1996) (Kamil et al. p. 313).
- "Simplifying" texts by removing structural cues and making shorter sentences can make texts harder to understand, because these techniques remove the links that provide parallel surface and conceptual meaning. The result is that coherence is reduced (Anderson, Hiebert, Scott, & Wilkinson, 1985; Beck & Dole, 1992; Beck & McKeown, 1989; Beck, McKeown, & Gromoll, 1989; Davison, 1984) (Kamil et al., p. 318).
- Understanding the performance of an individual often requires attending to differences in performance across a variety of reading activities including: variation in performance across types of text; the effect of non-print information; variation in performance across a variety of types of discourse and genres, including hypertext; the capacity to explore issues that go outside the traditional rubric of comprehension, such

as scanning, intertextuality, domain-specific strategies, and consulting illustrations. (RAND, pp. 57-58).

Best Practices in Literacy Assessment

Purposes for Assessment
Two major purposes for assessments are to inform instruction and to reflect the effect of instruction or intervention. Thus, an effective assessment system should provide not only important information about a child's relative standing in appropriate normative populations (school, state, and national norms/groups), but also important information about the child's relative strengths and weaknesses for purposes of educational planning. These assessments should be able to identify individual children as poor comprehenders, not only in terms of prerequisite skills, such as fluency in word identification and decoding, but also in terms of cognitive deficits and gaps in relevant knowledge (background, domain specific, etc.) that might adversely affect reading and comprehension, even in children who have adequate word-level skills. It is also critically important that such a system be able to identify early any child who is apt to encounter difficulties in reading comprehension because of limited resources to carry out one or another operation involved in comprehension (RAND, pp. 56-57).

Reading Disabilities
Children identified early on as poor readers remain poor readers; but targeted early intervention programs can greatly ameliorate these outcomes (Foorman & Torgesen, 2001). Consequently, there has been a flurry of activity to identify characteristics of effective intervention programs (Wasburn-Moses, p. 70). In the past thirty years, numerous studies have estimated the percent of students to have reading disabilities. A Canadian study (Commission on Emotional and Learning Disorders in Children, 1970) estimated that between 10% and 16% of school-age children required diagnostic and remedial help in reading. These findings were consistent with U.S. studies that 14.8% of grades 3 and 4 students (Mykelbust & Boshes, 1969) and 14% of students in grades 7-11 (Meier, 1971) met criteria for underachievement. More recently, a Connecticut longitudinal study (Shaywitz & Shaywitz, 1996) found that 17.5% of the population of school children in primary and middle school has reading difficulties (Snow, Burns, & Griffin, pp. 90-91).

Given that there will always be a percentage of students expected to have some reading difficulties, the use of early intervention (well-designed, focused instruction) is compelling. Hiebert et al. (1992) suggest that approximately 75% of children in the bottom quartile can be brought into range of their average peers when taught in groups of three, while Taylor, et al. (1992) provide evidence that approximately two thirds of this population can attain this level with focused instruction in groups of six or seven (Kamil, Mosenthal, Pearson, & Barr, p. 476). *Approximately the lowest 10% of the entire population appears to require even further support beyond these early interventions* (Kamil, Mosenthal, Pearson, & Barr, p. 477).

Risk Factors
Many other factors also affect literacy learning and development. Demographic data also suggest several risk factors of a majority of children with reading problem to include: being

from poor and/or stressed families, families with little education, families with a history of reading difficulties, homes lacking a supportive environment for literacy development and verbal interaction, and/or a home language other than English (Snow, Burns, & Griffin, pp. 119-131). In addition to outcomes assessments, assessments designed to reflect readers' cognitive, motivational, and linguistic resources as they approach a reading activity are necessary, because they can reflect the dynamic nature of comprehension (e.g., assessing increments of knowledge about vocabulary and particular target domains that result from interaction with particular texts). When the outcomes assessment identifies children who are performing below par, process assessments could help indicate why their reading comprehension is poor. Further, diagnostic assessments are crucial in dissecting the effect of particular instructional or intervention practices (RAND, p. 55).

Teacher Perceptions and Use of Assessment Data

Many literacy researchers have commented on the perceptions teachers have about use of standardized tests. Pressley, who has conducted numerous studies looking at reading instruction and the attributes of excellent teachers states that although excellent teachers give curriculum-based tests (e.g., spelling, fluency), there is little concern with standardized tests among these teachers. Many of the best teachers he's studied have expressed great reservations about the validity of standardized assessments forced on them (Johnson & Rogers, 2001).

Although many politicians argue that standardized testing will guarantee that poor and minority students receive a quality education, teachers have reported that testing pressures are affecting the quality of their instruction and their professional beliefs about reading and learning to read (Hoffman, Assaf, & Paris, 2001; Pennington, 2004). Many teachers may becoming less responsive and adaptive to students' literacy needs, while focusing more on skills management based on tested objectives (Flores & Clark, 2003) because of increased concerns about passing standardized tests (Assaf, p. 158).

Conversely, teachers see great value in informal assessment (formative) and do it constantly. They do not see that standardized assessments do much good for the children they teach (Pressley, pp. 355-356). Successful teachers use reading assessments for many purposes. They may use informal assessments to assess fluency, skills tests to diagnose strengths and weaknesses, observations of decoding and comprehension strategies during daily reading, and examine student work samples (Taylor & Pearson, p. 142). Thus, using large-scale assessment data to improve instruction, support student learning, and evaluate reading programs for effectiveness will not happen at the local levels, unless these perceptions change.

Hargreaves and Fink (2003) remind us that "High-stakes testing can push teachers to deliver improved results, but it does not necessarily cause them to produce better learning" (p. 695). We must remind ourselves of priorities, in terms of instructional time and the goals of deeper learning. The true goal of assessment is accomplished every day in schools where teachers systematically use assessment to inform their teaching. This kind of assessment rests on careful analysis of the strategies and knowledge that are required and on students' own strengths (Pinnell, p. 80).

Critical Outcomes and Potential Consequences for Literacy Assessment

Currently, widely used comprehension assessments are heavily focused on only a few tasks: reading for immediate recall, reading for the gist of the meaning, and reading to infer or disambiguate word meaning (RAND, p. 54). Assessment procedures to evaluate learners' capacities to modify old or build new knowledge structures, to use information acquired while reading to solve a problem, to evaluate texts on particular criteria, or to become absorbed in reading and develop affective or aesthetic responses to text have occasionally been developed for particular research programs, but have not influenced standard assessment practices.

Comprehension assessments that are narrow in focus may inadvertently limit the reading curriculum to preparation for those few tasks assessed. Most high stakes tests represent an over simplistic view of reading and have a narrow focus on particular reading skills and strategies (RAND, P. 54; Afflerbach, 2004, p. 7-10). Knowledge, application, and engagement are all critical outcomes of reading *with comprehension*; assessments that reflect all three of these outcomes are needed. Further, the absence of attention to these consequences in widely-used reading assessments diminishes the emphasis on them in instructional practices as well (RAND, p. 54). The Rand Reading Study Group proposes an approach that places more value on the usefulness of assessment procedures for improving instruction. Assessments need to be developed that are an integral part of and supportive of instruction.

When schools devote time and effort to testing and preparing students to take tests, they spend considerable resources to do so. This money might be spent in other ways, including enriched reading curriculum and teachers' professional development related to assessment (Stiggins, 2002). High stakes tests are expensive to purchase, to prepare for, to administer, and to score. Initiatives to support testing take from other, worthy initiatives related to fostering students' reading development (Afflerbach, 2004, pp. 7-10).

Research Related to Large-Scale Assessment

With the steady increase of accountability testing since 1980, there has been a parallel increase in concerns expressed about the liabilities of increased testing (Shepard, 2000). Some worry that the curriculum has been narrowed (Haertel, 1989); some worry that teachers are being judged inappropriately on the bases of standardized tests (Smith, 1991); and some worry that increased testing has negative effects on students' learning and motivation (Paris, Turner, Lawton, & Roth, 1991; Paris, 2000). The issue has such profound political and educational implications for reading that the International Reading Association (1999) and the American Educational Research Association (2000) published position papers pointing out the potential problems with high-stakes testing (Taylor & Pearson, p. 143). These concerns have been echoed in the more recent National Reading Conference Policy Brief,

With the focus now on accountability, rather than the diagnosis of learning for instructional purposes, (Campbell, 2002), several drawbacks to this shift in focus have been documented. Decisions about promotion or retention are attached to student success or failure on some state-mandated tests (Traub, 2002). Many teachers change their literacy curricula in order to train students to take the test (Harman, 2000). This change – from teaching for learning to teaching for the test – results in narrowing of the curriculum, loss of instructional time, and

loss of teacher autonomy (Campbell, 2002; Vacca & Vacca, 2001) (Higgins, Miller, & Wegmann, pp. 310-311).

Although it is not technically defensible for states to administer formal standardized measures to hold first and second graders to grade-level standards, policymakers have a legitimate concern that third grade is too late to identify children who are falling behind. Policymakers could reasonably require schools/districts to have procedures in place to monitor student progress using instructionally relevant assessments and a plan for providing intensified special help for children having difficulty, especially in learning to read (Shepard, Kagan, & Wurtz, p. 31).

- The intended use of an assessment – its purpose – determines the content (What should be measured?); the method of data collection (Should procedures be standardized?); technical requirements of the assessment (What level of reliability and validity need to be established?); and finally the stakes, or consequences of the assessment (Shepard, Kagan, & Wurtz, p. 6). For teaching and learning purposes, the timing of assessments makes the most sense if they occur on an ongoing basis as particular skills and content are learned. An assessment system may include assessment for these different purposes: to support learning, to identify special needs, to evaluate programs and monitor trends, and for high stakes accountability. Only under special circumstances would it be possible to serve more than one purpose with the same assessment (Shepard, Kagan, & Wurtz, p. 7).
- Many people believe high stakes tests are fair and scientific (Afflerbach, 2002). The vast majority of commercial and statewide reading tests are the result of considerable time and effort invested in developing and piloting the tests. Through adherence to what are for most people abstract notions of validity and reliability, tests can create a "scientific" aura. In reality, no research has been conducted that demonstrates a cause and effect relationship between increased high stakes testing and improvement in reading achievement scores (Afflerbach, 2004, pp 4-6).
- Tests have the ability to reduce and summarize complexities of reading to single raw scores and percentile rankings, and in doing so they appear almost magical (Lemann, 1999). The very high stakes tests that are believed to be scientific are actually severely limited in their ability to describe the wide range of reading achievement that most states and school districts set forth in their formal learning goals and standards statements (Davis, 1998). Understanding short texts and answering questions about them, required of all students taking high stakes tests, is but a small slice of what we expect accomplished student readers to do. In this sense, high stakes tests are an exceedingly thin measure of reading achievement and reading ability (Afflerbach, 2004, pp 4-6).
- The high-stakes nature of accountability assessments contributes to their possible inaccuracy. All assessments are fallible and potentially corruptible. Results can be distorted by departures from standardized administration procedures, or inappropriate teaching-to-the-test. These practices are documented to occur more frequently when the results have high-stakes consequences for students and teachers (Shepard, Kagan, & Wurtz, p. 31).

- High stakes tests most often come with caveats related to the accuracy of scores they produce and the suitability of uses of scores, and these caveats are widely ignored. Commercially produced reading tests and those created for statewide and federal high-stakes decision making regularly feature strong guidance related to the appropriate uses and misuses of test scores (Hamilton, Stecher, & Klein, 2002). Among the most frequent caveats is the admonition not to use a single high-stakes reading test score to make educational decisions. This caveat is based on the understanding that single test scores represent only one measure of student readers and that test scores are subject to natural variation and sampling error, as with the political polls that are conducted regularly in the United States. This means that a student's single high-stakes reading test score falls within a range of scores that represent the students' actual achievement level. When high stakes decisions are made using such unstable scores, the decisions may be faulty and costly. Psychometricians and test developers are fully aware of the dangers of such practice, yet it continues unabated (Afflerbach, pp. 11-15).
- High stakes tests are limited in their ability to describe students' reading achievement. A high stakes test score represents a single sample of what a student reader does on a standardized test. This score is not at all representative of the full range of reading that marks accomplished student readers. The single score provides only a snapshot of student performance. High stakes tests may well under represent reading performance, because these tests have a severely limited ability to describe complex reading and reading-related performances that mark the accomplished teaching and learning of reading (Afflerbach, 2004, pp 7-10).
- The technical standards for reliability and validity are much more stringent for high-stakes accountability assessment than for classroom assessments. The consequences of accountability assessments require that the instruments be sufficiency accurate to ensure that important decisions being made are not a result of measurement error (Shepard, Kagan, & Wurtz, p. 7).
- Increasingly, the failure to meet the standards is being associated with child-specific sanctions, such as retaining the child in grade or withholding a high school diploma. The achievement tests to which these high stakes are attached often reflect reading comprehension ability, even when the specific goal of the test is to assess knowledge in the content areas. The data available to date about the effect of high-stakes tests on student achievement are insufficient and conflicting. No research has addressed how poor comprehenders are selectively affected, either by the tests themselves or by the various consequences associated with them (RAND, p. 7).
- We know that retention in grade (an increasingly frequent consequence of failure on high-stakes assessments) does not improve long-term reading achievement without specialized instruction (RAND, p. 9).
- High stakes tests are used with increasing frequency to characterize and label young children who are in early developmental stages of reading. Children's growth and experiences related to reading vary widely prior to formal schooling, as does their experience with formal testing situations. We expect this varied experience to influence the skills, strategies, motivation, and conceptions of reading that young children bring to school. In contrast to this variety of experiences and abilities, high stakes tests force the labeling of young children and assignment to differential instruction that might not be appropriate or effective. Additionally, it is doubtful that

most young children have extensive standardized testing experience, so the very act of placing these children in such a situation introduces factors of familiarity and anxiety as possible influences on test performance (National Association for the Education of Young Children, 2003).

- Traditional modes of monitoring development of reading are standardized and norm-referenced instruments and criterion-referenced tests. Although these measures show where an individual falls within a peer group, they do not necessarily show in detail what an individual can do as a reader. *Some drawbacks of traditional modes include*: unreliable for making judgments about individual reading development; rarely have much demonstrated validity, as they assess only a narrow range of literacy activity; given infrequently, so even if reliable and valid, are of little use in planning and instruction; tend to narrow the curriculum, as teachers feel the need to "teach to the test;" and they can play a role in discouraging those children whose performance on the tests suggests that their reading development lags behind that of their peers (Allington & Walmsley, 1995, pp. 78-79; cf. Darling-Hammond, 1991; Stallman & Pearson, 1990). Additionally, traditional tests' content and format tend to focus instruction on coverage of the right answers (Meier, 2003) rather than on substantive learning (Braunger & Lewis, p. 131).
- Regarding the comprehension tests available for assessing reading, the RAND Reading Study Group (2002) asserts that they "sort children on a single dimension using a single method" (p.53) and lack a clear viable theory of comprehension. RAND calls for improvements in comprehension assessment that will accurately reflect the dynamic, developmental nature of comprehension and the interactions among reader, activity, text, and context (Braunger & Lewis, p. 131).
- Educators ought to be working toward the adoption of instructionally supportive accountability tests, designed to detect the kind of instructional impact that must be present if achievement gaps are ever to be *demonstrably* reduced. In 2001, the Commission on Instructionally Supportive Assessment described three attributes of an instructionally supportive achievement test: measure only a modest number of curricular aims; describe aims clearly so teachers can direct instruction at the curricular goals and not the test items; and supply score reports that show whether or not each curricular goal was met, so teachers can assess the effectiveness of instruction (Popham, p. 50).
- From the highest levels, policymakers in the state and federal government might use information from the National Assessment of Educational Progress (NAEP) to judge the effectiveness of reading instruction in an entire state or nation. The results provide a rough estimate of whether the policies are improving achievement. Large-scale assessments (norm and criterion-reference tests) are rarely sufficiently accurate by themselves, for individual diagnosis. At best, large-scale assessments can be used to screen children for further testing (e.g., using running records, IRIs, etc.) to confirm the results, or to diagnose specific reading problems (McKenna & Stahl, p.35).
- According to the National Institute of Child Health and Human Development (NICHD), too many children in our nation are having difficulty reading. The National Institute of Health (NIH), part of NICHD, has conducted and supported research addressing reading failure since 1965. They report that the illiteracy rates are unacceptably high. Over 40 percent of fourth grade students performed below basic

levels on the National Assessment of Educational Progress (NAEP) in 1994 and again in 1998. Over 10% of fourth grade children could not even participate in the NAEP because of severe reading difficulties.

- One way the results of NAEP are reported is by use of benchmarks. The NAEP Governing Board has divided performance into three levels – basic, proficient, and advanced. These performance levels are translated into specific benchmarks for each grade level. NAEP benchmarks were designed as high standards for children to reach, so that teachers would push students towards these standards, rather than a more modest level of achievement. "Children can fail to reach the basic level of NAEP at grade 4, but still demonstrate a literal level of understanding of what is read, understand a main idea of expository text, or follow a simple plot" (Donahue, Voelkl, Campbell, & Mazzeo, 1999). While we need high standards to propel our students to higher achievement, we are concerned that statements such as '40% of fourth graders are reading below the basic level' can be misinterpreted (McKenna & Stahl, pp.32-35).
- A test that reflects the curriculum that is taught is said to possess ***content*** *validity*. Alignment is used to ensure a good match between the standards that are taught and standards that are assessed. When a test produces results that conform well to real-world applications, it is said to possess ***construct*** *validity.* For example, if reading achievement results show that students who did well are already placed in advanced reading groups and students who performed poorly were already placed in below-average groups, then the test is said to have construct validity (McKenna & Stahl, pp. 35-36). Both content and construct validity are essential in designing test items for large-scale assessments.
- A shortcut to assessing general comprehension ability is to examine a student's ability to pronounce words, presented in graded lists of increasing complexity. Based on how far the student can progress, an estimate is made about the student's instructional reading level. This raises questions of *construct validity* – estimating a student's reading level using an assessment that does not require the student to read and involves no actual measure of reading comprehension – does not seem to make much sense. On the other hand, research does support correlations between word-list performance and more conventional approaches to comprehension. This method is best thought of as a "short cut" to estimate overall reading proficiency, but is no substitute for the "real thing" (McKenna & Stahl, pp. 174-175).
- "Without question, the worst norm-referenced score typically reported on group achievement tests is the grade-equivalent score. This norm purports to indicate the grade level at which the student performed on a particular test. A common way to interpret a grade-equivalent score is to relate it to that of "average children" at a particular level. For example, if a seventh grader produces a grade equivalent score of 2.6 on a test of reading comprehension, it is common for teachers to conclude that the student comprehends as well as the average second grader in the 6th month of school. This interpretation is difficult to defend in that it is unlikely that an assessment appropriate for middle school students would have been given to second graders to establish norms. Most grade-equivalent scores are merely statistical projections that permit a student to be compared with students of other ages." Additionally, grade-equivalent scores have "floors" and "ceilings" – the lowest and highest projections

that can be made. The International Reading Association Board (1980) officially condemned the use of grade-equivalent scores for these reasons (McKenna & Stahl, p.30).

Research Related to Assessment of Early Literacy

Although rarely the result of conscious policy decisions, a variety of indirect pressures – such as older kindergarteners, extensive preschooling for children from affluent homes, parental demands for teaching reading in kindergarten, and accountability testing in higher grades – have produced a skill-driven kindergarten curriculum. The result of these changes was an aversive learning environment inconsistent with the learning needs of young children. This in turn has lead to several ill-considered policies such as raising the entrance age for school, readiness screening to hold some children out of school, increasing kindergarten retentions, and 2-year programs for an extra year either before or after kindergarten. A review of controlled studies has shown no academic benefits from retention in kindergarten or extra year programs. Stallman and Pearson (1990) have shown the decomposed and decontextualized prereading skills measured by traditional readiness tests are not compatible with current research on early literacy and also raise serious equity concerns (Shepard, 1994, pp. 207-208).

The National Association for the Education of Young Children (NAEYC) has identified three legitimate purposes for assessment of young children: (1) to plan instruction and communicate with parents; (2) to identify children with special needs; and (3) to evaluate programs. In other words, tests should not be used if they do not bring about benefits for children (Shepard, 1994, p. 208). Gathering accurate information from young children is difficult and potentially stressful. Formal assessments may also be costly and take resources that could be used directly to provide programs and services. To warrant conducting assessments, there must be a clear benefit - either in direct services to the child or improved quality of educational programs (Shepard, Kagan, & Wurtz, p. 5).

Guiding principles for assessment of young children, consistent with NAEYC's perspective on testing are: *the content of assessments should reflect and model progress towards learning goals*, taking into account physical and social/emotional development; *methods of assessment must be appropriate to the development and experiences of young children*, meaning that in addition to written products, observation, oral reading, and interviews should be used for assessment, recognizing the diversity of learners; and *assessments should be tailored to a specific purpose*, assuring the validity of the assessment (Shepard, 1994, p. 208). Assessments designed for one purpose are not necessarily valid if used for other purposes. In the past, many of the abuses of testing with young children have occurred because of misuse (Shepard, Kagan, & Wurtz, p. 5).

Developing a complete, valid, and useful literacy assessment program requires more than choosing a new test or adopting a commercially packaged procedure. It requires consideration of the complex nature of literacy development and the principles of assessment that are derived from that complexity (Tierney, 1998).

Many assessment authorities have enumerated a variety of principles, Pennsylvania (*PA Literacy Framework*, 2000b, pp. 7.11-7.13) especially recommend the following five principles:

- **Literacy assessment should explicitly reflect the literacy goals and the experiences that lead to those goals** (McTighe, 1995)**.**
- **Literacy assessment should reflect an understanding that reading and writing are multi-dimensional, integrated, and revealed in performances over time** (Farr, 1992; McTighe, 1995)**.** State-mandated literacy assessments should be one component of a district's assessment package. As Tierney (1998) states, "Assessment should be viewed as ongoing and suggestive, rather than fixed or definitive" (p. 385). Long-term literacy engagement and interest requires sustained assessment efforts. An assessment package should contain instruments and procedures that allow the stakeholders to see growth over time.
- **Literacy assessment should reflect effective instructional practices** (Farr, 1992; McTighe, 1995). Assessment is not an end in itself. It is the process of collecting data in order to evaluate the effectiveness of the literacy practices being implemented in the classroom. Changes in testing have not kept pace with our knowledge of literacy development. A reliable and valid assessment package must contain accurate reflections of daily literacy performances.
- **Assessment procedures may need to be non-standardized to be fair to the individual** (Tierney, 1998). Though many tools within an assessment package will be standardized, there are situations (cultural differences, learning differences, etc.) that will make standardization impossible. As Tierney (1998) so aptly states, "Diversity should be embraced, not slighted" (p. 385). There will always be a tension between the need for uniformity and the need for measures that are sensitive to differences. A comprehensive assessment package balances standardized measures and non-standardized assessments. Effective classroom teaching does not occur by ignoring or removing diversities. The same is true for the design of assessments. Assessing learning within the context of diversity is the goal, and it is essential.
- **Assessment procedures should be child-centered, and they should support student ownership** (Farr, 1992; Tierney, 1998). Literacy assessment practices should be something the classroom teacher does with a learner, rather than something that is done to the learner. The ultimate goal of literacy instruction and assessment is to develop "habitual self-assessors" (Farr, 1998, p. 31).

In 1999, The International Reading Association wrote a position paper opposing high stakes literacy testing. That paper concludes on a positive note, however, by providing the following recommendations to teachers: Construct more systematic and rigorous assessments for classrooms, so that external audiences will gain confidence in the measures that are being used and their inherent value to inform decisions (*PA Literacy Framework*, 2000b, p. 7.33).

Other research findings on early literacy assessment…

- In line with a renewal of interest in early schooling and a belief in its critical role in later reading development, researchers documented the literacy development of 20 preschool children in five different locations, and then followed them into the first year of school. They found substantial variation in the reading capabilities of children entering school. Among many relevant findings, they showed that many children have knowledge about books, letters, and how to attend to print before entering school, but the first year of schooling is associated with significant gains in word concepts, punctuation, sentence writing, and a critical awareness that reading requires decoding. The most important finding is a caution about the "possible danger" with testing programs being used too early or interpreted as evidence of "risk" when in fact

children may simply have not had opportunities to learn what is being tested (Kamil, Mosenthal, Pearson, & Barr, pp. 5-6).

- One of the ironies of the screening measures used for detecting potential reading difficulties in young children is that the earlier a student takes these assessments, the less valid and potent a predictor the measure is. Most efforts to identify reading problems before receiving reading instruction over predicts reading disabilities (Jenkins & O'Connor, 2000). Thus, as we try to implement early intervention in kindergarten, we use measures with less precision than those for even slightly older children. By waiting a longer period, we will be more accurate, but will miss the opportunity of teaching virtually all children to read by the end of first grade (Gersten & Dimino, p. 104).
- It is difficult and perhaps impossible to design good measurement tools for use close to the onset of instruction. Standardized tests do not discriminate well until considerable progress has been made by children (Clay, p. 11). The *Observation Survey of Early Literacy Achievement* (developed by Marie Clay) is a set of standardized tasks that allow teachers to observe young children as they engage in reading and writing. Observation allows teachers to watch the child, to see at least part of the focus of his or her attention, to watch the child search for information in print and confirm what he/she thinks (Clay, p. 3). Systematic observations provide ways of knowing when we can make valid comparisons and decisions about the progress students are making in reading (Clay, p. 12).
- Systematic observation of children's reading and writing behaviors, under controlled conditions can provide valuable feedback for instruction and for program improvement. Such observation is an essential research tool in most scientific fields, and has proven to be an effective means for collecting objective information about the performance of young children on complex tasks (Genishi, 1982).
- The younger the child (especially before the age 6), the more difficult it is to obtain reliable and valid assessment data. Abstract paper-and-pencil tests may make it especially difficult for young children to show what they know (Shepard, Kagan, & Wurtz, pp. 5-6). Methods of collecting assessment data should include: direct observation in natural activities; looking at drawings and samples of work; asking questions, either orally or in writing; or asking informed adults about the child (Shepard, Kagan, & Wurtz, p. 12).
- During the development of Meisel's *Work Sampling System* (assessment portfolio), curriculum standards for young children were matched to observational criteria. If you observe only in a single instance and a single setting, you can be terribly misled. You have to have a multiplicity of data (systematic observation) from which to draw conclusions (Anzalone, McConnell, Barnard, Meisels, & Weinberg, pp. 4-5).
- Documenting the literacy development of young children can present many challenges. The most meaningful approach to assessment of individual young children is through continual observation by teachers and parents of children's progress in all developmental domains including social, emotional, physical, and cognitive. Performance inventories and portfolios of children's work provide a far more meaningful picture of the young child's progress than any standardized test results. Similarly, narrative reports by teachers, outlining children's progress, are far more useful at the primary level than numeric or letter grades, since they provide

information that can be used by parents to help their children at home (Katz, 1997). Katz reported that the main purposes for assessing children are: to determine progress toward literacy objectives; to make instructional or placement decisions; to pinpoint specific learning and teaching problems; to help inform instruction; to provide evidence of growth when reporting progress to parents and stakeholders; and to teach children how to self-assess their skills (*PA Literacy Framework*, pp. 2.27-2.28).

- Over the past decade, numerous research studies have linked phonological assessments to future success in reading. Scarborough (2005) cautions that the phonological model of predicting reading performance is not inaccurate, but it is incomplete. There are other non-phonological factors, such as proficiency in oral language, expressive vocabulary, and sentence or story recall that also can predict long-term reading outcomes (in grades 3-6). Scarborough's model accounts for students who decode well, but do not comprehend what they read. These students may have deficits in non-phonological areas. In the absence of appropriate screening and progress-monitoring assessments in this area, it is essential to integrate vocabulary and listening comprehension in the primary grades. Ongoing (daily) formative assessment will identify deficits in these areas and inform instruction and monitoring (Gersten & Dimino, p. 104).
- A disturbing trend we have observed in our field of research (Response to Intervention) is the treatment of many of the benchmarks on tests, such as DIBELS (Good & Kaminski, 2001) as if they descended from Mt. Olympus and were inviolate. Benchmarks are merely guidelines for indicating which students are *likely to be* reading at grade level or below grade level by the end of the year. As in medicine, where benchmarks are routinely used, we need to check that these numbers are stable over time for a given child, before taking serious action (Gersten & Dimino, pp 103-104).
- The intended use of an assessment will determine the need for normative information or other means to support the interpretation of assessment results. Identifying children with special needs requires normative data to distinguish serious physical, emotional, or learning problems from the wide range of normal development. When reporting to parents, teachers need some idea of what constitutes "grade-level performance" using norms in the form of grade-level benchmarks, rather than using statistical percentiles (Shepard, p. 209).
- Children need to know how patterns (not rules, such as "when 2 vowels go walking") help them to identify unknown words. Targeted assessment of word recognition can move from phonological awareness, to knowledge of sight words and decoding, to use of context, to fluent, automatic word recognition (McKenna & Stahl, pp.14-15).
- Knowledge of text structure can be assessed through free recall, which should roughly mirror the text structure (story elements of narrative or main idea and supporting details of expository text) (McKenna & Stahl, pp.18-19).
- *Fluency should be regularly assessed* - both formally and informally – to ensure that students are making progress. The most informal assessment is simply listening to a child read aloud and making a judgment about his/her fluency. Formal measures of fluency would look for a reading rate faster than 90 words per minute, reading with expression, and ability to comprehend what was read. Fluency assessments include using comparisons of timed samples with published norms, Informal Reading

Inventories (IRIs), miscue analyses, and running records (Armbruster & Osborn, p. 30).

- The evidence from research leads us to believe that miscue analysis can serve a useful but limited role in reading assessment. Three main implications for effective practice are recommended. (1) Use error totals (e.g., from running records) for the explicit purpose of determining independent and instructional reading levels. The use of semantically acceptable miscue tallies is not supported by research; (2) View meaning miscues (e.g., substituting horse for pony) as evidence of inadequate decoding skills, and focus teaching on use of graphophonic, semantic, and syntactic cueing systems; and (3) Study miscues to monitor progress toward relying more and more on decoding. Miscue analysis can do little to identify skills in word recognition, because those needs are not systematically assessed through oral reading and may be masked by reliance on context (Walpole & McKenna, 2006) (McKenna & Picard, 2006, p.379).
- Analyzing oral reading miscues can be beneficial, but perspectives on how it should inform instruction have changed. Ehri and McCormick (2004) suggested that one goal of instruction is to teach decoding, analogy, word prediction from context, and sight recognition as distinct strategies to identify words. Miscue analysis can help to monitor a child's ability to apply these strategies during reading. It can also reveal whether the balance is shifting over time away from context dependency towards automatic decoding (McKenna & Picard, 2006, p.380).
- IRI and running record assessments can help to determine a student's *independent* reading level (highest level of text a child can read without assistance), *instructional* level (highest level at which a child could benefit from instructional support), and *frustration* level (lowest level at which a child is likely to be frustrated, even with instructional support). These assessments generally begin administration at the student's estimated independent level and move to higher or lower text difficulty based on results. Coding of reading errors is used to analyze results for oral fluency and accuracy (word recognition). Comprehension scores can also be determined, but are not the primary focus of these assessments. Use of "leveled texts" for instruction (meaning texts leveled by complexity along a gradient that is sensitive to characteristics that reflect reading challenges) make the use of running record assessments valuable for tracking reading progress over time. For first graders, the first administration should be delayed until midyear so that more children will be capable of reading at least a pre-primer passage (McKenna & Stahl, pp.41-65).
- *Good phonics assessments are nearly always administered individually*, because the application of phonics skills requires that students produce pronunciations. Three common phonics assessments are the *Informal Phonics Survey* (assessing beginning consonants, 2-letter consonants, and vowel sounds using real words), the Z-test (a more advanced phonics test using "pseudo-words" all beginning with the/z/ sound and requiring students to decode patterns to produce the "rime" or end sound); and the *Test of Knowledge of Onsets* (teacher pronounces a word that ends the same and asks student to pronounce the new word with a different "onset," or beginning sound). Usefulness of these tests is not to produce a total score, but to identify specific strengths and weaknesses for instruction (McKenna & Stahl, pp.116-120).

- *Comprehension assessment* is somewhat a controversial topic, in that no general agreement exists on how to do it best. Traditionally, comprehension assessment has consisted of posing questions to answer at three levels of comprehension – *literal* (specific facts found explicitly in texts), *inferential* (having factual answers, as with literal questions, but requiring students to make logical connections among facts in texts), and *critical* (requiring students to use analysis or make value judgments about what was read) (McKenna & Stahl, pp.167-168).
- Young children do not often fully understand how stories work – especially complex stories with multiple problems. Researchers have found that young children often understand and remember only some parts of stories (Lipson, Mosenthal, & Mekkelsen, 1999). Grade 2 readers regularly recalled action and action-driven events when retelling stories, but tended not to include or recall in detail the internal responses (emotions) or internal plans (motivations linked to actions) of characters. To an extraordinary extent, when young children recall stories, the characters lack motivation, and it is often difficult to tell whether children have understood the causal links and/or tensions that mature readers expect from stories (Lipson et al., 1999). The retellings were however, always filled with action (Lipson & Cooper, pp. 2-3).
- The qualities of an effective and valid progress-monitoring tool for the early reading classroom or intervention program are similar to those of an effective diagnostic instrument, with the added requirements: (a) a sufficiently fine-grained scale to meaningfully measure change across time; and (b) the availability of equivalent multiple forms for each task to enable the teacher to re-administer the assessment at different points in time during the school year as a measure of growth (Fuchs & Fuchs, 1999) (Denton & Ciancio, p. 31).
- Regardless of whether an assessment is intended to measure early reading, knowledge of color names, or learning potential, assessment results are easily confounded by language proficiency, especially for children who come from homes with limited exposure to English (Shepard, Kagan, & Wurtz, p. 6).
- States considering early childhood assessments to monitor trends (low stakes assessments) could work to ensure that the content of classroom assessments used for promoting learning and development are closely aligned with content of the statewide assessment. These assessments would not be the same, but could be developed in parallel so they would be conceptually compatible and mutually supportive (Shepard, Kagan, & Wurtz, p. 34). Instructionally relevant assessments should reflect a clear continuum of progress in Grades K, 1, and 2 that lead to expected standards of performance by third and fourth grades (Shepard, Kagan, & Wurtz, p. 36).
- Beginning at age 5, it is possible to use direct measures, including measures of early learning, as part of a comprehensive early childhood assessment *for monitoring trends*. Matrix sampling (where students take different forms of the test with different items and there are no individual scores reported) should be used to ensure technical accuracy and to provide safeguards for individual children. Because of the cost of such an assessment, states should pick one grade level for monitoring trends in early childhood (e.g., kindergarten or grade 1) (Shepard, Kagan & Wutrz, p. 21). [Note: The NAEP test uses matrix sampling. No scores are reported for individual students and not all students at a grade level are tested.]

- Before age 8, standardized achievement measures are not sufficiently accurate to be used for high-stakes decisions about individual children and schools. Therefore, *high-stakes assessments intended for accountability purposes should be delayed* until the third grade (or preferably fourth grade) (Shepard, Kagan & Wutrz, p. 21).
- The most robust evidence about children's reading reveals developing skills that can be compared to individual standards of progress, as well as normative standards of achievement. A developmental approach balances the types of assessments across a range of reading factors and allows all stakeholders to understand strengths and weaknesses of the child's reading profile. Many teachers use this approach and we think it is a useful model for early reading assessment (Taylor & Pearson, p. 153).
- A developmental approach to early reading assessment is not a one-size-fits-all approach that gives the same test to all children on the same day. Instead, assessment is embedded in the daily classroom activities in which teachers use formal and informal assessment tools to ascertain if children are improving their literacy skills and knowledge, mastering the curriculum, and meeting community standards for literacy development. These practices are effective because they empower teachers and students alike (Taylor & Pearson, p. 159).
- Individually administered tests tend to provide more dependable results than group assessments because the teacher can command the mental engagement of the student, to some extent, during the testing process. On the other hand, group tests are far more efficient, even if their results are less dependable (McKenna & Stahl, p.23).
- More Kindergarten and grade 1 reading tests are individually administered today (52%) than in the early 1990s (18%), with about 20% of early literacy tests having the option of group or individual administration. Leveled books have become a common tool for determining reading levels (e.g., *DRA*) and developmental rubrics, teaching resources, and guidance for assessment techniques (e.g., *Meisels Work Sampling System*, *First Steps*) are available to support connections between assessment and instruction (Pearson & Meisels, CIERA presentation slides, 1999).
- Different types of assessments are used for different purposes, which include Instructional Planning (screening assessments, diagnostic assessments, and classroom-based assessments that monitor progress); and Program Evaluation (outcome assessments) (Torgesen et al., p. 6).
- The No Child Left Behind (NCLB) Act requires screening, diagnostic, and classroom-based instructional reading assessments to be administered in Grades K–3 (Title I, Part B, Subpart 1, Section 1202). The NCLB legislation (Title I, Part B, Subpart 1, Section 1208, Number 7) provides definitions of these assessments as follows (North Central Regional Educational Laboratory, p. 1; Torgesen et al., p. 14):
    - **SCREENING READING ASSESSMENTS** - Screening instruments should be relatively quick to administer, provide a gross indicator of "risk status" and must have good predictive utility for identifying the need for interventions. They are— valid, reliable, and based on scientifically-based reading research; and are *designed as a first step* in identifying children who may be at a high risk for delayed development or academic failure and in need of further diagnosis of their need for special services or additional reading instruction.
    - **DIAGNOSTIC READING ASSESSMENTS** - Diagnostic instruments must measure a variety of component skills or abilities, and must be directly useful in planning subsequent instruction. They are — valid, reliable, and based on

scientifically-based reading research; and *used for the purpose of— identifying a child's specific areas of strengths and weaknesses* so that the child has learned to read by the end of Grade 3. They are used to determine any difficulties that a child may have learning to read and the potential cause of such difficulties; and help to determine possible reading intervention strategies and related special needs.

- ❖ **CLASSROOM-BASED INSTRUCTIONAL READING ASSESSMENTS**– Progress monitoring instruments must have multiple forms, be quick and efficient to administer, and be *sensitive to growth over relatively short instructional intervals*. They may help identify students who need diagnostic assessment. They evaluate children's learning based on systematic observations by teachers of children performing academic tasks that are part of their daily classroom experience; and is used to improve instruction in reading, including classroom instruction.

- The Reading Assessment Committee for Reading First provides the following assessment guidelines for Reading Outcomes to be Assessed at Each Grade Level from K-3 (Torgesen et al., p.9):
  - ❖ **Kindergarten -** Phonemic Awareness, phonics, and vocabulary
  - ❖ **1st Grade -** Phonemic Awareness, phonics, fluency, reading comprehension, and vocabulary
  - ❖ **2nd Grade -** Fluency, reading comprehension, and vocabulary
  - ❖ **3rd Grade -** Fluency, reading comprehension, and vocabulary

How Some States Approach Early Literacy Assessment

Only a few states currently assess reading below grade 3 on a large scale (all of which are administered individually by teachers). Vermont has administered a customized version of the grade 2 *Developmental Reading Assessment (DRA*) since 1999. Louisiana and Rhode Island require the use of *DRA* for some (RI) or all students (LA). Virginia piloted in 2000 and now administers a criterion-referenced screening assessment, the *Phonological Awareness Literacy Screening (PALS)* to all Kindergarten through grade 3 students. *PALS* is touted to be a "state-of-the-art" large-scale literacy assessment. Some other states – Michigan, Texas, and Arkansas to name a few – have provided resources, professional development, and specific guidance to support early literacy assessment. (See also Appendix C for areas assessed by the specific tests listed.)

- **Arkansas**: Arkansas provides state-funded staff development that trains teachers in the administration and interpretation of two early reading assessment tools – Marie Clay's *Observation Survey* and the *Developmental Reading Assessment (DRA*). Teachers are taught how to use these assessments in accordance with the state requirements on early reading assessment, but are told that they can choose to use any other assessment they deem appropriate (Southwest Educational Development Lab. Reading Assessment Database).
- **Louisiana:** Legislation enacted during the 1997 Regular Legislative Session, directed that the State Board of Elementary and Secondary Education (SBESE), in cooperation with the State Department of Education (SDE) and all public schools, report on the number of students not reading on grade level in all first, second, and third grades throughout the state at the beginning of each school year. The data for each school and for each school system, and for the state as a whole, is required to be reported in the school progress profile

  In May 1998 the Board of Elementary and Secondary Education (SBESE) approved the *Developmental Reading Assessment* as the uniform assessment instrument to be implemented statewide in the fall of 1998. The *Developmental Reading Assessment (DRA),* an individually administered instrument, evaluates the two major aspects of reading: accuracy of oral reading and comprehension through reading and retelling of narrative stories - aspects of reading which are critical to independence as a reader. The assessment consists of twenty stories that increase in difficulty. Factors that contribute to the gradient of difficulty of the stories include the number of words on a page, complexity of vocabulary, story length, degree of support from the pictures, as well as complexity of sentence and story structure.

  From the fall of 1998 until the fall of 2003, the *DRA* was the uniform reading assessment used in Louisiana elementary schools to identify second- and third-grade children at risk of reading failure. During the fall of 2005, it continues to be used as the official reading assessment for many elementary schools. It is administered to second and third graders at the beginning of each school year and to first, second, and third graders at the end of each year. *DRA* results are reported by teachers on scannable forms. Data are compiled and reports are produced by the Louisiana State University Center of Assessment and

Evaluation. Reports provide the number of students assessed and identified as reading below grade level, on grade level, or above grade level.

While most school in LA continue to use the *DRA* as the mandated early reading assessment, other schools, at least one in each district, have received permission to use *DIBELS* in order to meet the guidelines for the Reading First grant. LA also developed the reading inventory, called the *Louisiana Literacy Profile (LLP*) to be used statewide; however, Reading First schools using the *DIBELS* no longer use the *LLP.* Many districts and teachers still use the *LLP* for informal classroom purposes or to assist with dyslexia screening in components not included in *DIBELS.*

The *Dynamic Indicators of Basic Early Literacy Skills* (*DIBELS*) are a set of standardized, individually administered measures of early literacy development. They are designed to be short (one minute) fluency measures used to regularly monitor the development of pre-reading and early reading skills. *DIBELS* provides teachers with benchmark scores that students need to meet at specific times during each year and suggests instructional interventions that are needed by students who fail to meet benchmark scores. Students who are in need of intervention may be divided into two categories: those at some risk of reading failure and in need of strategic intervention and those at greater risk and therefore in need of intensive intervention. *DIBELS* results are reported online and compiled by the University of Oregon (*Fall 2005 LA Reading Report).*

- **Michigan:** The battery of assessments included in the Kindergarten through Third Grade Michigan Literacy Progress Profile (MLPP, 2000) designed by the Michigan Department of Education and Michigan educators is intended to be resource for teachers to use selectively with some or all of their students at different points in time during grades K-3. The state legislature has recommended that the MLPP can be used to monitor annual student progress, as well as achievement in summer school programs. While there is some overlap of assessments used for Kindergarten through grade 3, the battery of assessments include: letter identification; letter-sound correspondence; phonological awareness; concepts about print; oral language and listening; decoding and word identification; oral reading fluency; journals and work samples; comprehension and retelling; and reading attitudes and habits (Taylor & Pearson, p. 155).
- **Texas:** Texas has created an assessment tool similar to Michigan's approach. The *Texas Primary Reading Inventory* (*TPRI*) is intended for teachers to use with Kindergarten through grade 2 children to monitor literacy progress (Taylor & Pearson, p. 155). The *TPRI*, currently used in 95% of Texas schools, is administered to all students whose scores fall below benchmarks on the screener, indicating that important reading concepts and skills are still developing. The purposes of the *TPRI* are to identify children at risk for reading problems in the early grades and to provide instructional information to teachers (Rathvon, pp. 284-285).
- **Vermont:** The *VT- DRA* (administered individually by trained classroom teachers and/or reading specialists) evaluates two major aspects of reading: accuracy of oral reading and text comprehension, through reading and retelling of narrative stories. Both aspects of reading are critical to independence as a reader. Oral readings are audio taped and a random sample is reviewed for assessor reliability and accuracy. Individual student scores indicate the child's

performance for ***both*** acceptable accuracy of oral reading and comprehension for the highest text level achieved (pre-primer through grade 3 and higher).

The *VT-DRA* has several potential benefits. First, it includes assessment tasks that closely align to the VT Grade Expectations (standards) being assessed. In addition, the tasks reflect high-quality, effective classroom practices: having children read stories at appropriate levels of difficulty, and observing closely their oral reading accuracy and comprehension. Such tasks should influence classroom practice in a positive manner, encouraging teachers to make sure that every student is reading books at appropriate levels of difficulty, and also promoting an increased focus on reading comprehension. Regularly using retellings emphasizes to students that reading involves both word recognition ***and*** thinking about what is read. The context of the *VT-DRA* - one to one conferences — is one that is appropriate for young children and provides the opportunity for the assessor to note a variety of skills, behaviors and strategies used (or not yet used) by the child. Finally, this assessment provides uniform information and benchmarks across schools. If children move from one school or district to another, the information will follow the child. The assessment should also provide consistency in expectations across schools by helping to set targets for key points in students' development as readers (VT Department of Education, 2005, pp. 3-5).

- **Virginia:** In Virginia, the *Phonological Awareness Literacy Screening (PALS*) is administered as a statewide reading assessment (fall and spring) for students in Kindergarten through grade 3. Reviewers have rated this an "outstanding example of the new generation of evidenced-based reading screening instruments" with high usability ratings and yielding a wealth of instructionally relevant information. "For a large-scale, teacher-administered battery, its technical quality is unsurpassed, with regular reviews and modifications to ensure the most accurate and valid measurement (Rathvon, pp. 250-259).
- **Rhode Island:** Because some school configurations in the state of Rhode Island include only grades K, K-1, or K-2, they are unable to administer the grade 3 large-scale assessment in reading, required by NCLB. For those schools, administration of the *Developmental Reading Assessment (DRA*) in Kindergarten and first grade is required. Classroom teachers administer the assessments and use an online service to enter student scores. *DRA* assessment kits are provided by the state and remain in the schools (Rhode Island Department of Education, pp. 3-4).

Appendix A

A Discussion of "Increasing Text Complexity"

Karin K. Hess and Sue Carey Biggam, 2004

[This article was produced in partnership with the New Hampshire, Rhode Island, and Vermont Departments of Education. Karin Hess is a Senior Associate with the National Center for the Improvement of Educational Assessment and Sue Biggam is the Associate Director of the Vermont Reads Institute at the University of Vermont. (Updated in 2006)]

The instruction and assessment of reading comprehension presents unique challenges to classroom teachers and test developers alike; and the criteria used in selecting a variety and range of appropriate texts are essential to meeting those purposes. In the classroom, students learn to apply and practice a variety of reading strategies, for different purposes and with different text types. Although short, contrived texts can be helpful for introducing a reading skill or strategy, students will not be able to develop effective comprehension strategies such as self-monitoring, summarizing, and self-questioning unless they are reading increasingly complex material of appropriately substantial length. Nor will they develop and acquire the rich vocabulary and broad understanding of text structure required to become a reader with excellent comprehension (Lipson & Cooper, p. 10).

Over time, students who are exposed to a variety of text types with increasing complexity also learn how text features differ by genre, and they gain confidence in pealing back the layers of complexity for a deeper understanding of what is read. In test development, the overall number of test items is driven by the length and type of reading passages and the number of items possible accompanying each text passage. Passages for reading assessment, drawn from "authentic" text whenever possible, should always include both literary and informational texts. A series of questions accompanying each reading passage may include initial understanding of text, analysis and interpretation of text, or a combination of both types of questions, especially for longer text passages.

We have learned from NAEP research (1985) that difficulty of text passages was one of the three most important factors in reading comprehension performance of 4th, 8th, and 12th grade students. The other two factors were familiarity with subject matter presented in text and the type (literal, inferential, etc.) of question asked (Chall and Conard, 1991). Other research suggests that at grades 2 and 3, word difficulty may influence text complexity more than other factors (Anderson, 1992). Lipson and Wixson (2003) summarize the challenges of understanding text complexity this way:

"In the past, one of the few text features that was given much attention was its difficulty or readability, as measured by factors such as the number of syllables in the words and the number of words in the sentences. Current research has demonstrated that a number of other factors have a significant impact on both how much and what students understand and learn from a text. The presence or absence of these factors determines the extent to which a given text can be considered 'considerate' (to enable readers with minimal effort) or 'inconsiderate' (text requiring much greater effort). (Armbruster, 1984) "

A variety of factors influence text complexity. The complexity of text, or the degree of challenge of a particular text, is the result of specific combinations and interactions of

these factors. For example, a text that has short simple sentences may, nevertheless, be challenging to read/comprehend when it contains abstract ideas, concepts that are unfamiliar, or requires a greater level of interpretation to unlock the intended meaning. Pinnell and Fountas' text leveling system (2002), an extension of the system used by Reading Recovery developed for classroom use at grades 3-6, includes these factors for determining complexity: understanding the nature of print, repeated text, natural language versus book text, supportive text, and high frequency vocabulary. Their system also calls attention to differences between fiction and nonfiction texts in book leveling, and includes descriptors that "overlap" to the next level of difficulty.

Chall, Bissex, Conard, and Harris-Sharples (1996) suggest that linguistic characteristics (vocabulary and sentence structure and variety) as well as concepts presented, text organization, and background knowledge required of readers all need to be considered in determining appropriateness of text for a given grade level. "Merely breaking up longer sentences and simplifying vocabulary does not guarantee that reading materials will be completely appropriate for lower reading levels." They also point out differences between popular fiction, literature, and informational texts with regard to text difficulty. For example, popular fiction tends to (a) use less figurative language than literature, (b) be more repetition of information, and (c) have more conventional language use; therefore demands on the reader of popular fiction are more about basic understanding of explicit messages than on interpretation of the message.

Criteria for increasing text complexity include factors that interact to affect the relative difficulty of reading particular material. The tables on the following pages describe specific ways in which text materials generally increase in difficulty over the grade span of grades 1 through high school. The descriptors in the tables build from one grade span to the next. It is expected that students would have experience reading text described for their grade levels, as well as those of earlier grade spans.

Factors that Interact to Influence Text Complexity

- **Word Difficulty and Language Structure,** including vocabulary and sentence type and complexity of words or structure (often determined through the use of multiple readability formulas)
- **Text Structure** (e.g., description, chronology, sequence/procedure, compare-contrast, cause-effect, proposition-support, problem-solution, critique, deductive/inductive); and coherence and unity devices used to clarify and connect ideas (e.g., pronouns, connectives, transitional devices, examples, ordered lists)
- **Discourse Style** (e.g., satire, humor)
- **Genre and Characteristic Features of the Text**
- **Background Knowledge and/or Degree of Familiarity with Content** needed by the reader (e.g., historical, geographical, or literary references)
- **Level of Reasoning Required** (e.g., sophistication or complexity of themes and ideas presented, abstract metaphors, etc.)
- **Format and Layout of Text,** including how text is organized, size and location (layout) of print, graphics, white space; and other book/print features (e.g., illustrations, headings, bold/italicized type, maps, charts, summaries)
- **Length of Text**

Complied for the Utah Department of Education, 2006
Updated 2007

Descriptors of Text Complexity for Grade Levels or Grade Spans

Karin K. Hess and Sue Carey Biggam, 2004

[Please Note: Sample grade-appropriate text titles are included at the end of the descriptors for each grade span as examples of text that would illustrate many of the characteristics described in the table. In many cases, particular teachers and schools will choose to introduce these specific texts at grade levels below or above the grade level indicated. While every descriptor might not be evident in a sample text passage, it is expected that the sample texts reflect the intent of the descriptors, and many of the indicators.]

Text Complexity Descriptors
End of Grade 1

- Includes a variety of literary texts (e.g., fantasy, realistic fiction, poetry), with some complexity in story structure (e.g., multiple episodes) and literary language.
- Simple informational books/text.
- Illustrations provide moderate support for the reader.
- Texts have several sentences per page, with sentences of moderate length and generally simple sentence structure.
- Very straightforward text structures (e.g., description, sequence).
- Familiar content.
- In narrative text, details related to story elements (setting, characters, events, resolution) provide strong support for both literal and interpretive meanings (e.g., for drawing basic inferences or basic conclusions).
- Informational texts use clear and consistent formats (e.g., print location on page), illustrations, and simple graphics to support understanding of content.
- Simple punctuation is used: period, question mark, exclamation point, quotation marks, and commas.

SAMPLE TEXTS AT THE END OF GRADE 1:

There's a Nightmare in my Closet; The Very Busy Spider; Nobody Listens to Andrew; Ants (Sunshine Science Series)

Text Complexity Descriptors
End of Grade 2

- Includes a variety of literary texts (e.g., realistic fiction, folktales, humorous stories, poetry) with elaborated episodes and events, and some extended descriptions.
- Stories usually have well-developed characters and episodes.
- Informational books/text.
- Some use of unfamiliar vocabulary, supported by other text features (e.g., headings, chapter titles, glossary).
- Illustrations may or may not be present on each page, but usually provide low to moderate support for the reader.
- Sentence structure becomes more complex – including causal phrases.
- Straightforward text structures in informational text (e.g., description, sequence, compare/contrast, problem/solution).
- Content usually familiar.
- In narrative text, details related to story elements (setting, characters, goals, attempts, consequences, and resolutions) provide moderate support for both literal and interpretive meanings (e.g., for predicting logical outcomes or drawing inferences about problem/solution).
- Informational texts use clear formats (e.g., use of simple headings to organize information into categories), illustrations that extend meaning, and simple graphics to support understanding of content.
- Full range of punctuation used, except dashes, colons, and semicolons.

SAMPLE TEXTS AT THE END OF GRADE 2

George and Martha; Cam Jansen and the Mystery of the Dinosaur Bones; The Stories Julian Tells; Happy Birthday Martin Luther King (Scholastic)

Text Complexity Descriptors
Grades 3-4

- ❖ Includes a range of longer literary selections, including realistic fiction and fantasies. Narratives usually include familiar characters or settings.
- ❖ Informational/functional text including short expository pieces (e.g., descriptive, compare/contrast, directions, simple recipes)
- ❖ Varied vocabulary, but generally familiar; some figurative language (e.g., similes). Increased use of challenging vocabulary (e.g., multi-syllabic words, words with multiple meanings). Technical words are defined, or explained in context.
- ❖ Sentence structure becoming more elaborate and complex, including some use of passive voice, abstract or descriptive language.
- ❖ Relatively straightforward text structures. Texts include more information, more complex ideas and relationships (e.g., examples, comparisons, cause/effect).
- ❖ Content usually builds from shared/somewhat familiar experiences.
- ❖ In narrative text, the story elements (plot, setting, characterization) provide support for both literal and interpretive meanings.
- ❖ Informational texts use clear formats, illustrations, and graphics to support understanding of content. Text features include timelines, captions, charts, and maps.
- ❖ Full range of punctuation used.

SAMPLE TEXTS AT THE END OF GRADE 3:	*SAMPLE TEXTS AT THE END OF GRADE 4:*
The Mouse and the Motorcycle; Sideways Stories; What's the Big Idea; Ben Franklin; Time for Kids magazine	*Cricket in Times Square; Castle in the Attic; WOW* magazine (National Wildlife Federation)

Text Complexity Descriptors
Grades 5-6

- ❖ Includes a range of literary selections, such as full-length novels, well-crafted short stories (with increasingly diverse characters and settings), historical fiction, and myths.
- ❖ Includes more complex informational/functional texts, such as persuasive essays, procedural "how to" guides, scientific and historical summaries (e.g., textbooks).
- ❖ More varied and challenging vocabulary, including use of figurative language (e.g., idioms, metaphors) and analogies. Some technical terms.
- ❖ Language in narrative text includes dialect and other linguistic variants to enhance characterization and setting.
- ❖ Ideas and content increase in number and density. Relationships between ideas become more complex (e.g. flashback may be introduced) in narrative text; graphs and charts are needed to convey key information in expository text.
- ❖ Content requires general background knowledge. Underlying themes become more complex and more universal.
- ❖ Interrelationships among story elements become more complex and require more interpretation. Literary elements include flashback, humor, suspense, personification, and exaggeration.
- ❖ Informational and functional texts use a variety of formats, illustrations, and graphics to support understanding. Text features include chapter headings, glossaries, punctuation guides.

SAMPLE TEXTS AT THE END OF GRADE 5:	*SAMPLE TEXTS AT THE END OF GRADE 6:*
Tuck Everlasting; Shh! We're Writing the Constitution; Cricket magazine	*True Confessions of Charlotte Doyle; Holes; The Grey King; Cobblestone* magazine

Text Complexity Descriptors
Grades 7 - High School

- ❖ Includes a full range of literary genres, including realistic and historical fiction, science fiction, fantasy, and folk literature.
- ❖ Informational/functional texts include primary sources, personal narratives and autobiographies, schedules, and manuals, as well as synthesized information found in textbooks.
- ❖ Increasing number of uncommon words, including words with non-literal meanings and more abstract vocabulary; word choice can reflect diverse historical and cultural context; text often includes technical words with specialized meanings.
- ❖ Language in narrative text is more elaborate and complex, and includes a wide range of dialogue, use of dialects, and varied sentence structure to convey specific meanings.
- ❖ Prose style matches text purpose (informational, recreational, provocative, etc.).
- ❖ Relationships between ideas become less explicit and require more inference or interpretation.
- ❖ Understanding content requires increasing cultural and historical breadth of knowledge.
- ❖ More sophisticated and complex themes.
- ❖ Texts often call for literary analysis.
- ❖ Informational texts use format, illustrations, and graphics to support understanding of meaning.
- ❖ Text features often include advance organizers, inset text, and technology support.

SAMPLE TEXTS AT GRADE 7:	*SAMPLE TEXTS AT GRADE 8:*	*SAMPLE TEXTS AT HIGH SCHOOL:*
Roll of Thunder, Hear My Cry; Diary of a Young Girl; Muse magazine	*The Upstairs Room; Narrative of the Life of Frederick Douglass; The Giver; Science* magazine	*To Kill a Mockingbird; Night; Into Thin Air; Newsweek* magazine

The following sources were referenced to develop text complexity descriptors:

- Armbruster and Osborn, *Put Reading First*: *The research building blocks for teaching children to read.* National Institute for Literacy, 2001.
- Caldwell, *Reading Assessment*, Guildford Press, 2002.
- Crafton, *Standards in Practice: Grades K-2*. Urbana, IL: National Council of Teachers of English, 1996.
- Chall, Bissex, et al. *Qualitative Assessment of Text Difficulty: A Practical Guide for Teachers and Writers*. Brookline Books, Inc., 1996.
- Fountas and Pinnell, *Guiding Readers and Writers Grades 3-6*. Heinemann, 2001.
- Lipson, M. & Cooper, J. D. Current research in reading/language arts: Understanding and supporting comprehension development in the elementary and middle grades. Houghton Mifflin Co, 2002.
- Lipson, M., Mosenthal, J., & Mekkelsen, J. The nature of comprehension among grade 2 children: Variability in retellings as a function of development, text, and task. In T. Shanahan & F. V. Rodriguez-Brown (Eds.), *Forty-eighth yearbook of the National Reading Conference* (pp. 104-119). Chicago, IL: National Reading Conference, 1999.
- Lipson and Wixson, *Assessment and Instruction of Reading and Writing Difficulties: An Interactive Approach*. Pearson Education, 2003.
- National Assessment Governing Board and NAEP Reading Consensus Project. *Draft NAEP Reading Framework*. US Government Printing Office, April 2002.
- National Assessment Governing Board and NAEP Reading Consensus Project. Reading Framework for the National Assessment of Educational Progress: 1992-2000. US Government Printing Office, 2000.
- New Standards Primary Literacy Committee. Reading and Writing Grade by Grade: Primary Literacy Standards for Kindergarten through Third Grade. National Center of Education and the Economy and the University of Pittsburgh, 1999.
- Pinnell and Fountas, Leveled Books for Readers Grades 3-6. Portsmouth, NH: Heinemann, 2002.
- Rees and Shortland-Jones with Education Department of Western Australia. Reading Developmental Continuum, Portsmouth, NH: Heinemann, 1994.
- Sierra-Perry. *Standards in Practice: Grades 3-5*. Urbana, IL: National Council of Teachers of English, 1996.
- US Department of Education, Reading: Knowing What Works, K-3. National Institute for Literacy, 2000.
- Weaver, Brenda. Leveling Books K-6: Matching Readers to Text. International Reading Association, Inc., 2000.
- Wilhelm. *Standards in Practice: Grades 6-8*. Urbana, IL: National Council of Teachers of English, 1996.

Appendix B

Reading Fluency Rates

Complied by Susan Carey Biggam and Karin K. Hess, 2004

[This information was originally complied in partnership with the New Hampshire, Rhode Island, and Vermont Departments of Education for the New England Common Assessment Program/NECAP. Karin Hess is a Senior Associate with the National Center for the Improvement of Educational Assessment (www.nciea.org) and Sue Biggam is the Associate Director of the Vermont Reads Institute at the University of Vermont. (Updated in 2006)]

Recommended Fluency Rates (in words read correctly per minute)									
	Grade K	**Grade 1**	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**
Oral:	**N/A**	**50-80**	**80-100**	**90-120**	**115-140**	**125-150**	**135-160**	**140-175**	**150-180**
Silent:	**N/A**	**N/A**	**N/A**	**115-140**	**130-175**	**160-200**	**190-220**	**215-245**	**235-270**

The following sources were referenced to determine recommended fluency rates:

- Armbruster & Osborn, *Put reading first*: *The research building blocks for teaching children to read*, National Institute for Literacy, 2001.
- Caldwell, *Reading assessment*, Guilford Press, 2002.
- Fountas and Pinnell, *Guiding readers and writers grades 3-6*, Heinemann, 2001.
- Lipson and Wixson, *Assessment and instruction of reading and writing difficulty*, Pearson Education, 2003.
- NAEP's Scale for Assessing Oral Reading Fluency, 2001.

Appendix C

An excellent way to track literacy progress grade-to-grade is to develop a literacy profile for each student, documenting ongoing assessment results across the 5 domains of reading (phonological awareness, phonics, fluency, comprehension, and vocabulary), as well as assessment of oral language development, concepts of print, reading strategies, reading attitudes/interests, and writing skills. There are a variety of measures that can be used to gather data for each area of early reading; different types of assessments are used for different purposes, which include Instructional Planning (screening assessments, diagnostic assessments, and classroom-based assessments that monitor progress); and Program Evaluation (outcome assessments). Below is a partial listing of early literacy assessments frequently being used by classroom teachers and schools

Menu of *Some* Frequently Used Literacy Assessments for Grades K-3
(Compiled by Karin K. Hess, Center for Assessment, 2006, updated 2007)

Name of Assessment	Potential Purposes of Assessment	What it Assesses	Comments
Observation Survey of Early Literacy Achievement	Screening Diagnostic Progress Monitoring	Concepts of Print Fluency Writing Vocabulary	Individually administered; Developed in New Zealand by Marie Clay, founder of Reading Recovery, an early intervention program
*Texas Primary Reading Inventory (TPRI)***	Screening	Phonological Awareness Listening Comp (K) Phonics Fluency Comprehension	Individually administered; ** Comprehensive state-sponsored early reading battery; Texas benchmarks for grades K-2 depending on subtest
Dynamic Indicators of Basic Early Literacy Skills (DIBELS)	Screening Progress Monitoring Outcome	Fluency	Individually administered; Read graded passages aloud; benchmarks for grades 1-3
Peabody Picture Vocabulary Test III (PPVT)	Screening Diagnostic	Listening Comprehension Vocabulary	Individually administered
Phonological Awareness Literacy Screening	Screening	Phonological Awareness Fluency Comprehension	Untimed and individually administered; State-of-the-art statewide reading screening battery; VA benchmarks for some subtests
*Comprehensive Test of Phonological Processing (CTOPP)**	Diagnostic	Phonological Awareness	Individually administered; identifies children significantly below peers in grades K-1, at risk for reading difficulty * Listed among Reading-First Recommended Diagnostic tests
*Early Reading Diagnostic Assessment (ERDA)**	Diagnostic	Listening Comprehension Oral Expression (Retell) Vocabulary	Individually administered; Used primarily for K & Grade 1 * Listed among Reading-First Recommended Diagnostic tests
*Test of Word Reading Efficiency (*TOWRE)*	Diagnostic Progress Monitoring	Fluency (K-2 scores tend to reflect accuracy rather than reading fluency)	Individually administered * Listed among Reading-First Recommended Diagnostic tests
*Woodcock-Johnson III**	Diagnostic	Listening Comprehension Phonological Awareness Comprehension	* Listed among Reading-First Recommended Diagnostic tests
Yopp-Singer Test of Phonemic Segmentation	Progress Monitoring	Phonological Awareness	Very difficult for most K children

Menu of *Some* Frequently Used Literacy Assessments for Grades K-3 (continued) **Classroom-Based Assessments (for informal/ongoing assessment and formal assessments)**			
Name of Assessment	**Potential Purposes of Assessment**	**What it Assesses**	**Comments**
Running records (with miscue analysis)	Progress Monitoring	Accuracy & Fluency Comprehension	Individually administered
Developmental Reading Assessment (DRA) Grades K-3	Progress Monitoring Outcome	Accuracy Comprehension	Individually administered; Used by VT as large-scale assessment gr 2; RI required early reading assessment K-1 (some schools); LA required for all gr 1-3 students
Rigby Benchmark Kits (Grades K-2)	Progress Monitoring Outcome	Accuracy Fluency Comprehension	Includes fiction and non-fiction texts, uses running records
Curriculum-based measurement (CBM)	Screening Progress Monitoring	Fluency Comprehension	Individually administered; Oral reading of passages drawn from graded material; Excellent for local norming; can be used K-6
Qualitative Reading Inventory III (QRI III)	Diagnostic Progress Monitoring	Accuracy Fluency Comprehension	Individually administered; usually given 2 times a year, comparing child to a benchmark
The Critical Reading Inventory (Pearson)	Diagnostic Progress Monitoring	Comprehension Habits of Reading	Use with grades 1-high school; analyze how readers use background knowledge and habits of thinking
Book logs; surveys; interviews/conferences; reading response journals; written responses to what is read, with scoring guides; systematic observation; think-aloud protocols; graphic organizers; work samples	Progress Monitoring	Comprehension Vocabulary Text Structure Knowledge Reading Attitudes Habits of Reading	Individual and/or group administration

Sources Referenced for Assessments

Applegate, M., Quinn, K., & Applegate, A. (2006). Profiles in reading comprehension. *The reading teacher,* 60, 48-57.

Balanced reading assessment [online] available: www.balancedreading.com/assessment.html

Cappellini, M. (2005). *Balancing reading and language learning: A Resource for teaching English language learners, K-5*. Portland, ME: Stenhouse Publishers.

Early reading assessment: A guiding tool for instruction [online] available: www.k8accesscenter.org/training_resources/earlyreadingtools.asp

Institute for the Development of Educational Achievement (2002-2004). Big ideas in beginning reading. University of Oregon [online] available http://readinguoregon.edu/assessment

McKenna, M. & Stahl, S. (2003). *Assessment for reading instruction*. NY: Guilford Press.

Rathvon, N. (2004). *Early reading assessment: A practitioner's handbook*. NY: Guilford Press.

Southwest Educational Development Lab. Reading assessment database for grades k-2. [online] available: www.sedl.org

Torgesen, J., Kame'enui, E., Francis, D., Fuchs, L. Good, R., O'Connor, R., Simmons, D., & Tindal, G. (2002). *Linkages in developing and implementing Reading First assessment plans.*

[online] available http://idea.uoregon.edu/assessment/sept_02_workshops/pres_four.pdf

Wren, S. (2004, November). Descriptions of early reading assessments. Southwest Educational Development Laboratory. [online] available http://www.balancedreading.com/assessment/assessment.pdf

Appendix D

Bibliography of Resources Cited

Afflerbach, P. (2002). The road to folly and redemption: Perspectives on the legitimacy of high-stakes testing. *Reading research quarterly*, *37*, 348-360.

Afflerbach, P. (2004). High Stakes Testing and Reading Assessment. National Reading Conference Policy Brief. Presentation at the University of Maryland September 2004.

Anzalone, M., McConnell, S., Barnard, K., Meisels, S., & Weinberg, R. (Spring 2000). Minnesota Round Table 1999: Observation and assessment of young children: Issues in research, policy, and practice. Conference sponsored by Center for Early Education and Development and University of Minnesota.

Applegate, M., Quinn, K., & Applegate, A. (2006). Profiles in reading comprehension. *The reading teacher,* 60, 48-57.

Armbruster, B. & Osborn, J. (2001). *Put reading first*: *The research building blocks for teaching children to read.* National Institute for Literacy [online] available: www.nifl.gov

Assaf, L. (2006). One reading specialist's response to high-stakes testing. *The reading teacher,* 60, 158-167.

Bishop, P., Reyes, C., & Pflaum, S. (2006). Read smarter, not harder: Global reading comprehension strategies. *The reading teacher,* 60, 66-69.

Block, C. & Pressley, M. (Eds.) (2002). *Comprehension instruction: Research-based best practices.* NY: Guilford Press.

Braunger, J. & Lewis, J. P. (2001). *Building a knowledge base in reading* (2nd ed). Newark, DE: International Reading Association and National Council of Teachers of English.

Clay, M. (2002). *An observation survey of early literacy achievement* (2nd ed.). Portsmouth, NH: Heinemann.

Cunningham, J., Spadorcia, S., Erickson, K., Koppenhaver, D., Sturm, J., & Yoder, D. (2005). Investigating the instructional supportiveness of leveled texts. *Reading research quarterly*, 40, 410-427.

Denton, C., Ciancio, D., & Fletcher, J. (2006). Validity, reliability, and utility of the Observation Survey of Early Literacy Achievement. *Reading research quarterly* 41, 8-34.

Dole, J., Duffy, G., Roehler, L. & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. *Review of educational research*, 61, 239-264.

Dorr, R. (2006). Something old is new again: Revisiting language experience. *The reading teacher,* 60, 138-146.

Dymock, Susan. (2005). Teaching expository text structure awareness. *The reading teacher,* 59, 177-181.

Eldredge, J., Reutzel, D., & Hollingsworth, P. (1996). Comparing the effectiveness of two oral reading practices: Round-robin reading and shared book experiences. *Journal of literary research*, 28, 201-225.

Fuchs, L. S. & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School psychology review*, 28, 659-671.

Genishi, C. (1982). Observational research methods for early childhood education. In B. Spodek, *Handbook of research in early childhood education*. New York: The Free Press.

Gersten, R. & Dimino, J. (2006). RIT (Response to Intervention): Rethinking special education for students with reading difficulties (yet again). *Reading research quarterly* 41, 99-108.

Hamilton, L, Stecher, B., & Klein, S. (2002*). Making sense of test-based accountability in education.* Washington, DC: RAND.

Heller, M. (2006). Telling stories and talking facts: First graders' engagement in a nonfiction book club. *The reading teacher,* 60, 358-369.

Higgins, B., Miller, M., & Wegmann, S. (2006). Teaching to the test…not! Balancing best practice and testing requirements in writing. *The reading teacher,* 60, 310-319.

International Reading Association. (adopted 1999). High-stakes assessment in reading: A position statement of the International Reading Association. Newark, DE: Author [online] available http://www.reading.org/pdf/high_stakes.pdf

Kamil, M., Mosenthal, P., Pearson, P.D., & Barr, R. (Eds.) (2000). *Handbook of reading research,* Vol. III. Mahwah, NJ: Lawrence Erlbaum, Associates.

Lipson, M. & Cooper, J. D, (2002). Current research in reading/language arts: Understanding and supporting comprehension development in the elementary and middle grades. Houghton Mifflin Co.

Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. New York: Farrar, Straus & Giroux.

Lipson, M., Mosenthal, J., & Mekkelsen, J. (1999). The nature of comprehension among grade 2 children: Variability in retellings as a function of development, text, and task. In T. Shanahan & F. V. Rodriguez-Brown (Eds.), *Forty-eighth yearbook of the National Reading Conference* (pp. 104-119). Chicago, IL: National Reading Conference.

Lousiana Department of Education (2005). *Fall 2005 Louisiana Reading Report*. LA: author.

McKenna, M. & Stahl, S. (2003). *Assessment for reading instruction*. NY: Guilford Press.

McKenna, M. & Picard, M. (2006). Revisiting the role of miscue analysis in effective teaching. *The reading teacher,* 60, 378-380.

Meyers, A. M. (2005/2006). The princess storyteller, Clara clarifier, Quincy questioner, and the wizard: Reciprocal teaching adapted for kindergarten students. *The reading teacher*, 59, 314-324.

Nagy, W. & Scott, J. (2000) Vocabulary processes. In Kamil, Mosenthal, Pearson, & Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 269- 284). Mahwah, NJ: Erlbaum.

National Association for the Education of Young Children. (2003). *Early childhood curriculum, child assessment and program evaluation: Building an accountable and effective system for children birth through age eight*. Author [online] available http://www.naeyc.org/resources/position_statements/positions_intro.asp

National Reading Panel. (2000). Report of the National Reading Panel: *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for Reading instruction.* Washington, D.C.: National Institute of Child Health and Human Development (NICHHD), National Institutes of Health [online] available: www.nationalreadingpanel.org

North Central Regional Educational Laboratory (2005). Starting Points for Meeting the Requirements for Screening, Diagnostic, and Classroom-Based Instructional Reading Assessments in the Reading First Program of the No Child Left Behind Act of 2001. Author [online] available http://www.ncrel.org/litweb/start/

Pearson, D., & Meisels, S. (Snowbird 1999). CIERA Session: Three perspectives on early reading assessment (presentation slides). [online] available: www.CIERA.org

Pennsylvania Department of Education. (2000a). Early literacy learning. In *Pennsylvania literacy framework* (Chapter 2). Harrisburg, PA: Author [online] available: http://www.pde.state.pa.us/reading_writing/lib/reading_writing/PaLiteracyFramework2000OpeningPages.pdf

Pennsylvania Department of Education. (2000b). Assessing literacy. In *Pennsylvania literacy framework* (Chapter 7). Harrisburg, PA: Author [online] available: http://www.pde.state.pa.us/reading_writing/lib/reading_writing/PaLiteracyFramework2000OpeningPages.pdf

Pinnell, G. (2006). Every child a reader: What one teacher can do. *The reading teacher,* 60, 78-83.

Popham, W. J. (November, 2004). A game without winners. *Educational leadership*, 62, 46-50.

Pressley, M. (2002). *Reading instruction that works: The case for balanced teaching.* NY: Guilford Press.

Purcell-Gates, V., Duke, N., & Martineau, J. (2007). Learning to read and write genre-specific text: Roles of authentic experience and explicit teaching. *Reading research quarterly*, 42, 8-45.

RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: author.

Rathvon, N. (2004). *Early reading assessment: A practitioner's handbook.* NY: Guilford Press.

Rhode Island Department of Education (February 2006). Rhode Island state assessment program: District and school testing coordinators handbook for K-1 program 2006. Providence, RI: Author

Samuels, S.J. (1979). The method of repeated readings. *The reading teacher*, 32, 403-408.

Saul, W. & Dieckman, D. (2005). Choosing and using information trade books. *Reading research quarterly*, 40, 502-513

Scharer, P., Pinnell, G. Lyons, C. & Fountas, I. (October, 2005). Becoming engaged readers. *Educational leadership,* 63, 24-29.

Snow, C., Burns, M.S., & Griffin, P. (Eds.) (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Shepard, L. (November, 1994). The challenges of assessing young children appropriately. *Phi delta kappan*, 206-212.

Shepard, L., Kagan, L., & Wurtz, E. (Eds.) (1998). *Principles and recommendations for early childhood assessment.* Washington, DC: National Education Goals Panel.
[online] available: http://www.negp.gov/webpg720.htm#child

Sipe, L. & McGuire, C. (2006). Young children's resistance to stories. *The reading teacher,* 60, 6-13.

Spencer, B. & Guillaume, A. (2006) Integrating curriculum thorough the learning cycle: Content-based reading and vocabulary instruction. *The reading teacher,* 60, 206-219.

Stahl, K. (2005). Improving the asphalt of reading instruction: A tribute to the work of Steven Stahl, *The reading teacher,* 59, 184-192.

Stiggins, R. (2002). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan, 83,* 758-765.

Taylor, B. & Pearson, P. D. (Eds.) (2002). *Teaching reading: Effective schools, accomplished teachers.* Mahwah, NJ: Lawrence Earlbaum Associates.

Torgenson, J. (1998). Catch them before they fall: Identification and assessment to prevent reading failure in young children. [online] available http://www.readingrockets.org/article.php?ID=411

Torgesen, J., Kame'enui, E., Francis, D., Fuchs, L. Good, R., O'Connor, R., Simmons, D., & Tindal, G. (2002). *Linkages in developing and implementing Reading First assessment plans.* [online] available http://idea.uoregon.edu/assessment/sept_02_workshops/pres_four.pdf

VT Department of Education (Revised October 2005). *Vermont Developmental Reading Assessment (VT-DRA): Administrator's manual*. Montpelier, VT: author.

Wasburn-Moses, L. (2006). 25 best Internet sources for teaching reading. *The reading teacher,* 60, 70-75.